Proof, Sets, and Logic

M. Randall Holmes

version of 4/2/2021

For Jonathan

Contents

0.1 Introduction 9
0.1.1 Version Notes 9
0.1.2 Introductory Remarks 12

1 Proof **13**
1.1 Basic Sentences 14
1.2 Conjunction 15
1.3 Disjunction 17
1.4 Implication 18
1.5 Biconditional and Exclusive Or 19
1.6 Negation and Indirect Proof 20
1.7 Generality and the Rule of Substitution 32
1.8 Note on Order of Operations 32
1.9 Quantifiers 33
1.9.1 Variation: Hilbert symbols and definite descriptions . . 35
1.10 Proving Quantified Statements and Using Quantified Hypotheses 36
1.10.1 Reasoning with the Hilbert symbol 38
1.11 Equality and Uniqueness 39
1.12 Dummy Variables and Substitution 40
1.13 Are we doing formal logic yet? 42
1.14 Exercises 43

2 Typed theory of sets **51**
2.1 Types in General 51
2.1.1 (digression) Unsorted Preamble 52
2.1.2 (digression) Typed language introduced 56
2.2 Typed Theory of Sets 58
2.3 Russell's Paradox? 64
2.4 Simple Ideas of Set Theory 66

2.4.1 Review of set theory proof strategies 73
2.4.2 Exercises 77
2.5 Digression: simple ideas of set theory in the language of the unsorted preamble 79
2.6 Finite Number; the Axiom of Infinity; Ordered Pairs 83
2.6.1 Digression: The Quine Ordered Pair 92
2.6.2 Exercises 97
2.7 Relations and Functions 104
2.7.1 Exercises 107
2.8 Digression: The logic of subjects and predicates, or second-order logic 108
2.9 Defining Functions by Recursion; First-Order Peano Arithmetic 112
2.9.1 Exercises 127
2.9.2 A case study: proof of the left distributive law in formal arithmetic 129
2.10 Equivalence Relations, Partitions, and Representatives: the Axiom of Choice 132
2.10.1 Exercises 136
2.11 Cardinal Number and Arithmetic 137
2.11.1 Exercises 145
2.12 Number Systems 147
2.12.1 Exercises 152
2.13 Well-Orderings and Ordinal Numbers 153
2.13.1 Exercises 160
2.14 Transfinite Induction and Recursion 162
2.14.1 Exercises 174
2.15 Lateral Functions and T operations; Type-Free Isomorphism Classes 175
2.15.1 Lateral functions in the system of the unsorted preamble: an axiom of embedding 178
2.16 Other Forms of the Axiom of Choice 179
2.16.1 Exercises 184
2.17 Transfinite Arithmetic of Order, Addition, and Multiplication 185
2.17.1 Exercises 189
2.18 Cantor's Theorem 190
2.18.1 Cardinal Exponentiation and the Theorem 190
2.18.2 Applications to the Number Systems 191

2.18.3 Cardinals and Ordinals; Cardinal Successor; The Hartogs and Sierpinski Theorems . . . 194
2.18.4 Hierarchies of Cardinals; A Disadvantage of Strong Extensionality . . . 197
2.19 Sets of Reals . . . 199
2.20 Complex Type Theories . . . 199
2.21 Infinite Indexed Families; König's Theorem . . . 199
2.22 Partition Relations . . . 199
2.23 Large Cardinals . . . 201
2.24 Pictures of Sets: the Theory of Isomorphism Types of Well Founded Extensional Relations . . . 202
2.24.1 Coding Sets in Relations . . . 202
2.24.2 Passing to Isomorphism Types . . . 207
2.24.3 The Hierarchy of Ranks of Set Pictures . . . 212
2.25 Category theory . . . 214

3 Untyped theory of sets 221
3.1 The original system of Zermelo . . . 223
3.1.1 Exercises . . . 228
3.2 Basic set constructions in Zermelo set theory . . . 229
3.2.1 Relations and Functions . . . 231
3.2.2 Exercises . . . 236
3.3 Case study: the implementation of the number systems in untyped set theory . . . 238
3.3.1 The natural numbers . . . 238
3.3.2 The natural numbers and counting elements of sets . . . 248
3.3.3 The positive rationals . . . 257
3.3.4 Magnitudes (positive reals) . . . 258
3.3.5 The real number system . . . 258
3.4 Preliminaries for transfinite arithmetic of cardinals and ordinals . . . 260
3.5 Zorn's Lemma, The Well-Ordering Theorem, and the official definition of cardinality . . . 270
3.6 The cumulative hierarchy picture and Replacement . . . 273
3.6.1 Basic definitions of ordinal and cardinal numbers in untyped set theory; the cumulative hierarchy introduced 273
3.6.2 More on the von Neumann definitions of ordinal and cardinal number . . . 279
3.6.3 The Axiom of Replacement and *ZFC* . . . 283

3.6.4 Transfinite Induction and Recursion 291
3.7 The theory of infinite ordinal and cardinal numbers in untyped set theory 295
3.7.1 Transfinite ordinal arithmetic 295
3.7.2 Transfinite cardinal arithmetic 297
3.7.3 Exercises 308
3.8 Logically regimented set constructions and the definition of L 310
3.8.1 Defining the well-ordering on L 316
3.8.2 L satisfies the axioms of ZFC 320
3.8.3 L satisfies GCH 323
3.8.4 Final remarks about L 325
3.9 Theories with proper classes 327
3.9.1 Pocket set theory, or, who said mathematicians don't have a sense of humor? 328
3.10 Forcing 332
3.10.1 Independence of CH 346
3.11 Independence of Choice 349
3.12 † Bridges from untyped set theory to typed set theory 354
3.12.1 †The intended interpretation of Zermelo set theory in set pictures; the Axiom of Rank; transitive closures and Foundation 354
3.12.2 †Digression: Interpreting typed set theory as Mac Lane set theory 359
3.12.3 †Translation between Type Theory and Set Theory . . 362

4 Logic **367**
4.1 Formalization of Syntax and Substitution 367
4.2 A toy example (calculator arithmetic) 368
4.3 A formal syntax for our type theory 373
4.3.1 Exercises 377
4.4 Formalization of Reference and Satisfaction 378
4.4.1 Exercises 383
4.5 Formal Propositional Sequent Calculus 384
4.6 Formal First-Order Sequent Calculus: the Completeness, Compactness and Löwenheim-Skolem Theorems 388
4.6.1 Exercises 397
4.7 Cut Elimination for First-Order Logic 398
4.8 Incompleteness and Undefinability of Truth 399

5 Model Theory **403**
5.1 Ultrafilters and Ultrapowers 403
5.2 Technical Methods for Consistency and Independence Proofs . 407
5.2.1 Frankel-Mostowski Methods; The Independence of Choice 407
5.2.2 Constructibility and the Minimal Model of Type Theory 407
5.2.3 Forcing and the Independence of CH 408
5.2.4 Generalizing the T operation 408
5.2.5 Forcing: Basic Definitions 409

6 Saving the Universe: Stratified Set Theories **413**
6.1 Introducing *NFU* 413
6.1.1 Typical Ambiguity Examined 414
6.1.2 Definition and Consistency of *NFU* 417
6.1.3 Mathematics in *NFU* 421
6.1.4 There are Urelements 424
6.2 Extensions of *NFU* 424
6.2.1 The Axiom of Counting; ω-Models. 424
6.2.2 The Axiom of Cantorian Sets; the Axiom of Large Ordinals 424
6.2.3 The Axiom of Small Ordinals; the BEST model 424
6.3 The Extensional Subsystems 424
6.3.1 Ambiguity in Theories with Finitely Many Types; NF_3 424
6.3.2 Predicativity; NFP; The Ramified Theory of Types Interpreted in NFP; NFI 428
6.4 Finite Universes: *NFU* + "the universe is finite". 428
6.5 New Foundations 428
6.5.1 History of *NF*; Errors of Quine 428
6.6 Technical Methods for Consistency and Independence Proofs in *NF(U)* 429
6.6.1 Forcing in Type Theory and Set Theory 429
6.6.2 Frankel-Mostowski Permutation Methods 429
6.7 Cut Elimination in Type Theory and Set Theory 429
6.8 Stratified Combinatory Logic and λ-Calculus 429
6.9 Rieger-Bernays Permutation Methods 429
6.10 Limitations of Universal Constructions 429

7 Philosophy of Set Theory **431**

8 Appendix: Manual of Logical Style 433
8.1 Introduction . . . 433
8.2 Conjunction . . . 434
8.2.1 Proving a conjunction . . . 434
8.2.2 Using a conjunction . . . 434
8.3 Implication . . . 435
8.3.1 Proving an implication . . . 435
8.3.2 Using an implication . . . 435
8.4 Absurdity . . . 436
8.4.1 Proving the absurd . . . 437
8.4.2 Using the absurd . . . 437
8.5 Negation . . . 437
8.5.1 Proving a negation . . . 437
8.5.2 Using a negation: . . . 438
8.6 Disjunction . . . 438
8.6.1 Proving a disjunction . . . 439
8.6.2 Using a disjunction . . . 439
8.7 Biconditional . . . 440
8.7.1 Proving biconditionals . . . 441
8.7.2 Using biconditionals . . . 441
8.8 Calculating with biconditionals . . . 441
8.9 Universal Quantifier . . . 442
8.9.1 Proving Universally Quantified Statements . . . 442
8.9.2 Using Universally Quantified Statements . . . 443
8.10 Existential Quantifier . . . 444
8.10.1 Proving Existentially Quantified Statements . . . 444
8.10.2 Using Existentially Quantified Statements . . . 445
8.11 Proof Format . . . 445
8.12 Examples . . . 446

9 Appendix: Description of the logic of Marcel 453

0.1 Introduction

0.1.1 Version Notes

These are notes to myself as the editor of the document. I will highlight changes which actually affect material currently being lectured (or past material), which will of course also be of interest to current students.

4/2/2021: C. Ryan-Smith pointed out some corrections needed in section 2.1.1. Thank you!

3/30/2021: minor edits

3/28/2021: More edits, adding proof of the left distributive law in formal arithmetic.

3/26/2021: complete rewrite and rearrangement of the suggestion of an unsorted foundation for the type theory of chapter 2.

3/2/2021: (some of these dhanges already posted on the first): new material on rules for reasoning about sets, and proof strategies for reasoning about sets, and a new homework set.

2/11/201: some revised and additional solutions. Allowed a dummy line $a = a$ at the beginning of a UG block (actually, reflexivity of equality already allows this line!)

2/9/2021: logic homework updates

2/4/2021: Added another problem on p. 42

1/13/2021: Added a problem to the exercises on p. 42.

1/11/2021: Just signalling the new term of Math 402/502 at this point. Changes in the text will be signalled if made.

3/24/2019: Corrected an error in the forcing section which I am reading and may soon revise. Inserted the alternative presentation of type theory as an unsorted theory.

12/6/2017: The proof of independence of CH by forcing is added; I hadn't realized that I hadn't written it in the notes.

There is an indication of a considerable simplification of my forcing treatment, by restricting the construction of names, which appears to eliminate the atom problem without reintroducing the mutual recursion issue.

12/5/2017: Notes on independence of choice from ZFA added. Some exercises of an unsatisfactory nature are given: some of them are things you might be able to do.

12/4/17: Fixed stupid typo in an exercise. Repaired definition of D^{∞}.

12/1/2017: Fixed error in the definition of x_p in the forcing section. Expanded on the discussion under one of the problems.

11/29/2017: Notes on theories with proper classes added.

11/16/2017: There are now exercise sets at the ends of sections 3.8 and 3.10. I have no confidence in these, but one must ask something!

11/15/2017: Removed the first approach to collapsing names and cleaned up typos and editing disasters found during lecture. Also added headings for paragraphs so that it is easier to see the structure of the text.

11/14/2017: Further cleanup of the second approach to collapsing names. There were considerable difficulties (including a lot of cut and paste errors) now, one hopes, fixed. It was in an awful state: apologies to anyone who tried to read it.

11/13/2017: I wrote out the alternative approach to collapsing names in full. I think it is preferable, and with a little extra notation not noticeably harder to follow. It has the advantage that we talk about equality of collapsed names rather than a further equivalence relation on collapsed names. 6:30 pm debugging the proof.

11/10/2017: I have the construction of the forcing model *correctly* described (the description I gave in lecture was incorrect). I'm less charmed with this approach than I was when I thought it worked the way I described it in lecture, but it *does* work. 3:30 pm typo 10 am on the 11th, typo fixes and a footnote detailing an alternative approach.

11/9/2017: Considerable tightening of the initial part of the forcing section, up to the definition of name closures and an indication of the reasons that the name closure construction works.

11/7/2017: slight tweak while preparing forcing slides.

11/3/2017: Notes on forcing added. These are still rough, but ought to see the light of day... Removed section 5.1, which was implemented as section 3.8. 4 pm more proofreading of section 3.10.

11/2/2017: aiming to install the L lectures. The L lectures (at least a first draft) are now installed. Comments are seriously invited!

11/1/2017: several minor edits

10/24/2017: Added rough notes up to the global well-ordering on L.

10/21/2017: Added the note that Separation follows from Replacement plus the existence of the empty set, in the section where Replacement is introduced. I added a proof of the Mostowski Collapsing Lemma at the end of the same section.

10/18/2017: Typo fixes (thank you, students). Fixed up the language which provides for predicates picking out every element of D in the logically regimented sets construction: 0 codes sethood and $\{x\}$ codes the predicate which picks out just x.

10/16/2017: corrected a typo.

10/13/2017: Enhancements to the section on logically regimented constructions inspired by today's lecture. It was not nearly as much in error as I thought; I added the stipulation that there is a predicate associated with each element of D, and I added the formal definition of the constructible universe.

10/11/2017: Adding exercises.

10/10/2017: Adding infinite sums and products of cardinals and König's theorem.

10/6/2017: Added enhancements from the lecture of today. The incorrect statement of the definition of cofinality in lecture is corrected; I claim insomnia!

10/4/17: correction in proof of $\kappa + \kappa = \kappa$. This and other improvements in corresponding sections in both chapter 2 and chapter 3.

10/2/17: corrected embarrassing misstatement of Zorn's Lemma in recent material. Thanks to student :-)

10/1/2017: Adding material on transfinite arithmetic of cardinals and ordinals in chapter 3. This required a couple of corrections to the chapter 2 material as well.

9/29/2017: 12pm: Updated the discussion of the cumulative hierarchy with the definition of hierarchy along a well-ordering given in class. I'm still working on defining the homework assignment. 3 pm: a few more revisions based on today's lecture, and homework problems added in sections 3.5 and some parts of 3.6.

9/28/2017: More work on logically regimented constructions. I need to think about possibly reconciling the current section 3.8 and chapter 5 eventually.

9/27/2017: drafting a section on representing sets associated with general logical formulas. I'll be using this soon, but I think it will need more updates.

9/25/2017: tightened up the text in all of chapter 3. Added a new section on transfinite induction and recursion for a pending lecture.

9/23: added lecture on Zorn's Lemma and the definition of cardinality.

9/22: minor typo fix. 4:30 pm: finished notes for the Wednesday lecture (added to section 3.4); notes for the Friday lecture will be coming shortly as a new section 3.5.

9/19/17, 9/20/17: Fixed some typos interchanging $\oplus$ and $\otimes$ in the discussion of counting sets in the untyped set theory section. Another typo of the same kind fixed on the 20th.

9/16/2017, 6 pm, 7 pm: Most notes for the lecture of the 15th are written. Exercises should appear sometime later this evening or tomorrow. 7 pm exercises on p. 238-9 ready.

9/15/2017, 5 pm: I have completed the section on counting finite sets, with the discussion of the definitions of cardinal addition and multiplication. I still need to write material covering the lecture on the 15th, and only after that will I post exercises.

9/6/2017: Expansions and corrections motivated by the 9/6 lecture.

9/5/2017: Working on section 3.3.2 on the natural numbers and counting elements of finite sets.

9/1/2017: Added more text to section 3.3.1 on implementing the natural numbers in set theory, with exercises. 3 pm version suggested an alternative approach to one of the problems which does not work (as I found on trying to write it out); the 3:30 version corrects this.

8/31/2017: Adding notes about arithmetic from the 8/30 lecture.

8/27/2017: Corrected some terminology in section 2.24 on category theory to standard form, in response to a student comment.

8/25/2017: Changed from article format to book format. Revisions mostly of section 3.2.1 inspired by class presentation of this section (fixing chapter 2 language there). Added section 3.2.2 of exercises.

8/24/2017: Very minor layout changes. 232

8/23/2017: Revisions on first day of class. Moved daggered subsections of chapter 3 to a separate section on bridges between untyped and typed set theory at the end.

8/22/2017: Duplicating some of the development in the typed set theory chapter in the untyped set theory chapter to support Math 522 instruction, notably terminology related to relations and functions.

7/10/2017: Added a discussion of bounded separation in the list of Zermelo axioms. Started writing the section on logically motivated set constructions in chapter 5.

7/6/2017: Various edits through the existing text. Started working on the index. Added some empty sections which will need to be filled in for my Math 522 intentions and some comments on already existing sections which are empty or partial which will need development for Math 522. One point is that I'll be doing chapter 5 work in untyped set theory which I originally intended to do in type theory: what I can do is present it in a way that it is clear how to do it in both approaches.

7/5/2017: Added treatment of the Hilbert symbol and definite description operator in the first part, with footnotes on how to treat it in type theory if it is adopted. 6:30 pm working on refinements of the untyped set theory chapter.

6/1/2017: I'm doing wildly speculative things in the category theory section.

5/31/2017: Clearing the version notes at the end of the Spring 2017 class taught from these notes. 1:30 pm edits up to the beginning of the section on number systems. I'm planning to review and extend the document this summer. 3:30 pm a few more edits here and there.

0.1.2 Introductory Remarks

This is being written as a textbook for Math 502, Logic and Set Theory, and Math 522, Advanced Set Theory, at Boise State University, on the practical level. [1]

On the Platonic level, this is intended to communicate something about proof, sets, and logic. It is about the foundations of mathematics, a subject which results when mathematicians examine the subject matter and the practice of their own subject very carefully.

The "proof" part refers to an informal discussion of the practice of mathematical reasoning (not all *that* informal) which will serve as a springboard for the "logic" component. It also introduces formal notation for logic.

The "sets" part refers to a careful development of mathematical ontology (a fancy word for "what are we talking about"?): familiar mathematical concepts are analyzed in terms of the fundamental concepts of *set* and *ordered pair*. This chapter gives us an opportunity to practice the proof skills of which chapter 1 provides an overview. A distinctive feature of our development is that we first develop basic concepts of set theory in a typed theory of sets, then make the transition to the more usual untyped set theory in a separate chapter.

The "logic" part refers to a much more formal discussion of how we prove things, which requires both the "proof" and "sets" components to work properly, and in which bits of language (sentences and noun phrases) and proofs are actually mathematical objects.

All of this is supported by some software: the formal logic introduced in chapter 4 (and one of the alternative set theories introduced in chapter 6) are the logic of our sequent theorem prover Marcel, to which we will have occasion to refer, and which will be used for some lab exercises. We hope to find that experience with Marcel will assist the learning of formal logic.

The final chapter on alternative set theories will probably not be reached in the course (or in a first course, at any rate) but has some bearing on other ways we could get from type theory to set theory and on the way set theory is implemented in Marcel.

[1] currently being used for Math 522 for the first time in Fall 2017, which will require addition of much new material!

Chapter 1

Proof

In this chapter we discuss how we make "formal proofs" (really, as we will see in the Logic chapter, rather *informal* proofs) in English, augmented with formal notation.

Our framework is this. We will identify basic logical structures of statements. Statements have two fundamental roles in proofs which need to be carefully distinguished: there are statements which we are trying to deduce from our current assumptions, which we will call "goals", and there are statements already assumed or deduced from the current assumptions which we are allowed to use, which we will call "posits". The reason we call these last "posits" instead of something like "theorems" or "conclusions" is that posits may be consequences of statements which we have only assumed for the sake of argument: a posit is not necessarily a theorem. For each basic logical structure, we will indicate strategies for deducing a goal of that form (from the currently given posits) and strategies for using a posit of that logical form to deduce further consequences. Further, we will supply formal notation for each of the basic logical structures, and we will say something about the quite different English forms which statements of the same underlying logical form may take.

It is useful to note that my use of the word "posit" is eccentric; this is not standard terminology. We can adopt as a posit any current assumption, any previously proved theorem, or anything which follows logically from current assumptions and theorems. We allow use of "posit" as a verb: when we adopt A as a posit, we posit A (to posit is either to *assume* for the sake of argument or to *deduce* from previous posits).

We are trying to say carefully "deduce" rather than "prove" most of

the time: what we can *prove* is what we can deduce without making any assumptions for the sake of argument.

1.1 Basic Sentences

Sentences in mathematical English (being sentences of natural language) have subjects, verbs and objects. Sentences in formal mathematical language have similar characteristics. A typical mathematical sentence already familiar to you is $x < y$ (though we will see below that we will usually call this particular (grammatical) sentence a "formula" and not a "sentence" when we are being technical). Here x and y are noun phrases (the use of letters in mathematical notation is most analogous to the use of *pronouns* in English, except that for precision of reference mathematical language has a lot more of them). $<$ is the verb, in this case a transitive verb with subject and object. In the parlance of mathematical logic, a transitive verb is called a "binary predicate". Another typical kind of mathematical sentence is "x is prime". Here the verb phrase "is prime" is viewed as an intransitive verb (we don't distinguish between adjectives and intransitive verbs as English does). We can't think of examples of the use of intransitive verbs in mathematical English, though we are sure that they do exist. An adjective or intransitive verb is a "unary predicate" in mathematical logic. Two commonly used words in mathematical logic which have grammatical meanings are "term" and "formula": a "term" is a noun phrase (for the moment, the only terms we have are variables, but more term constructions will be introduced as we go on) and a "formula" is a sentence in the grammatical sense ("sentence" in mathematical logic is usually reserved for formulas not containing essential references to variables: so for example $x < y$ is a formula and not (in the technical sense) a sentence, because its meaning depends on the reference chosen for x and y, while $2 < 3$ is a formula and a sentence (no variables) and $(\exists x.x < 2)$ is a formula and a sentence (the x is a dummy variable here)). What we call "basic sentences" (using terminology from grammar) in the title of this section will really be called "atomic formulas" hereinafter.

The English word "is" is tricky. In addition to its purely formal use in "x is prime", converting an adjective to a verb phrase, it is also used as a genuine transitive verb in formulas like "x is the square of y", written $x = y^2$ in mathematical language. The $=$ of equality is a transitive verb (as far as we are concerned: it is not treated the same by English grammar) and also

part of our basic logical machinery.

The English word "is" may signal the presence of another binary predicate. A formula like "x is a real number" may translate to $x \in \mathbb{R}$, where $\in$ is the predicate of *membership* and $\mathbb{R}$ is the name of the set of all real numbers. For that matter, the formula "x is prime" could be read $x \in \mathbb{P}$ where $\mathbb{P}$ is here supposed to be the set of all prime numbers.

In our formal language, we use lower case letters as variables (pronouns). There will be much more on the care and feeding of variables later on. Some special names for specific objects will be introduced as we go on (and in some contexts lower case letters (usually from the beginning of the alphabet) may be understood as names (constants)). Capital letters will be used for predicates. $P(x)$ ("x is P") is the form of the unary predicate formula. $x\,R\,y$ is the form of the binary predicate formula. Predicates of higher arity could be considered but are not actually needed[1]: a ternary predicate formula might be written $P(x, y, z)$. The specific binary predicates of equality and membership are provided: $x = y$, $x \in y$ are sample formulas. Much more will be heard of these predicates later.

We will have another use for capital letters, mostly if not entirely in this Proof part: we will also use them as variables standing for sentences. We use variables A, B, C for completely arbitrary sentences (which may in fact have complex internal structure). We use variables P, Q, R for propositions with no internal structure (atomic formulas). Once we get to the chapters on set theory we will once again allow the use of capital letters as variables representing objects (usually sets): the grammar of our language will prevent confusion between capital letters used as terms and capital letters used as unary or binary predicates.

1.2 Conjunction

This brief section will review the mathematical uses of the simple English word "and". The use of "and" as a conjunction to link sentences is what is considered here. If S is "snow is white" and G is "grass is green", we all know what "snow is white and grass is green" means, and we formally write

[1]The precise point here is that we do not require ternary predicates if we have a notion of ordered pair, as $T(x, y, z)$ (T a hypothetical ternary predicate) can be understood as abbreviating $T(x, \langle y, z\rangle)$, and predicates with four or more arguments can be reduced to binary predicates similarly.

$S \wedge G$.

Certain English uses of "and" are excluded. The use of "and" to link noun phrases as in "John and Mary like chocolate" is not supported in mathematical language. This use does have a close connection to the logical "and": the sentence is equivalent to "John likes chocolate and Mary likes chocolate". One should further be warned that there is a further complex of uses of "and": "John and Mary went out together" does not admit the logical analysis just given, nor (probably) does "John and Mary moved the half-ton safe". There is an example of the nonlogical use of "and" in mathematical parlance: there is a strong tempation to say that the union of two sets a and b, $a \cup b$, consists of "the elements of a and the elements of b". But $x \in a \cup b$ is true just in case $x \in a$ *or* $x \in b$. Another example of a use of "and" which is not a use of $\wedge$ is found in "x and y are relatively prime".

We note and will use the common mathematical convention whereby $t\,R\,u\,S\,v$ is read $t\,R\,u \wedge u\,S\,v$, as in common expressions like $x = y = z$ or $2 < 3 \leq 4$. This chaining can be iterated:

$$t_0\,R_1\,t_1\,R_1\,t_2\,\ldots\,t_{n-1}\,R_n\,t_n$$

can be read

$$t_0\,R_1\,t_1 \wedge t_1\,R_2\,t_2 \wedge \ldots \wedge t_{n-1}\,R_n\,t_n.$$

Proof Strategy: To deduce a goal of the form $A \wedge B$, first deduce the goal A, then deduce the goal B.

This rule can be presented as a rule of inference

$$\frac{\begin{array}{l}A\\B\end{array}}{A \wedge B}$$

We call this rule *conjunction introduction* (or just *conjunction*) if a name is needed.

If you have posited (assumed or deduced from current assumptions) $A \wedge B$, then you may deduce A and you may deduce B.

This can be summarized in two rules of inference:

$$\frac{A \wedge B}{A}$$

$$\frac{A \wedge B}{B}$$

We call this rule *simplification* if a name is needed.

The operation on propositions represented by $\wedge$ is called *conjunction*: this is related to but should not be confused with the grammatical use of "conjunction" for all members of the part of speech to which "and" belongs.

1.3 Disjunction

This subsection is about the English word "or".

Again, we only consider "or" in its role as a conjunction linking sentences; the use of "or" in English to construct noun phrases has no analogue in our formal language.

When we say "A or B" in mathematics, we mean that A is true or B is true *or both*. Here we draw a distinction between senses of the word "or" which is also made formally by lawyers: our mathematical "or" is the "and/or" of legal documents. The (presumably) exclusive or of "You may have chocolate ice cream or you may have vanilla ice cream" is also a logical operation of some interest but it is not yet introduced here.

We write "A or B" as $A \vee B$, where A and B are sentences.

Proof Strategy: To deduce a goal $A \vee B$, deduce A. To deduce a goal $A \vee B$, deduce B. These are two different strategies.

This can also be presented as a rule of inference, which comes in two different versions.

$$\frac{A}{A \vee B}$$

$$\frac{B}{A \vee B}$$

The rule is called *addition* if a name is needed.

We will see below that two more powerful strategies exist (generalizing these two): To deduce a goal $A \vee B$, assume $\neg A$ ("not A") and deduce B; To deduce a goal $A \vee B$, assume $\neg B$ and deduce A. We call both of these rules *disjunction introduction* (or *alternative elimination*).[2]

For a fuller discussion of this kind of proof strategy which involves the introduction of an additional assumption, see the subsection on implication below (and for more about negation see the section on negation below).

To use a posit $A \vee B$ (assumed or deduced from the current assumptions) to deduce a conclusion G, we use the strategy of *proof by cases*: first deduce G from the current assumptions with A replacing $A \vee B$, then deduce G from the current assumptions with B replacing $A \vee B$ [both of these proofs are needed].

The operation on propositions represented by $\vee$ is called *disjunction*.

1.4 Implication

The sentences "if A, then B", "B if A", "(that) A (is true) implies (that) B (is true)" all stand for the same logical construction. Other, specifically mathematical forms of the same construction are "(that) A (is true) is sufficient for B (to be true)" and "(that) B (is true) is necessary for A (to be true)". We provide optional padding phrases in parentheses which are needed in formal English because a proposition cannot grammatically live in the place of a noun phrase in an English sentence. Our formal notation for any of these is $A \to B$.

Don't spend a lot of time worrying about "necessary" vs. "sufficient" for purposes of reading this text – I only occasionally use them. But other writers use them more often; if you are going to read a lot of mathematics you need to know this vocabulary.

It is important to notice that unlike previous (and subsequent) constructions this one is not symmetrical: "if A, then B" is not equivalent to "if B, then A".

[2]It is a common error (or redundancy at least) to present proofs of a disjunction by alternative elimination in both forms, by a false analogy with the method of proof by cases.

Proof Strategy: To deduce a goal $A \to B$, assume A (along with any other assumptions or previously deduced results already given in the context) and deduce the goal B. Once the goal B is proved, one withdraws the assumption that A and all consequences deduced from it (it is local to this part of the proof). The same remarks apply to the negative assumptions introduced in the rule of alternative elimination for proving disjunctions indicated above.

We call this rule *deduction*.

An alternative strategy for proving $A \to B$ (called "indirect proof" or "proving the contrapositive") is justified in the section on negation: assume $\neg B$ and adopt $\neg A$ as the new goal.

A posit of the form $A \to B$ is used together with other posits: if we have posited $A \to B$ and we have also posited A, we can deduce B (this rule has the classical name *modus ponens*). We will see below that we can use posits $A \to B$ and $\neg B$ to deduce $\neg A$ as well (the rule of *modus tollens*).

Another way to think of this: if we have a posit $A \to B$ we can then introduce a new goal A, and once this goal is proved we can deduce the further conclusion B. [or, following the pattern of *modus tollens*, we can introduce a new goal $\neg B$, and once this goal is proved we can deduce $\neg A$].

The operation on propositions represented by $\to$ is called *implication*.

The additional strategies indicated in this section and the section on disjunction which involve negation ($\neg$) will be further discussed in the section on negation below.

1.5 Biconditional and Exclusive Or

When we say "A if and only if B", "A (being true) is equivalent to B (being true)", "A exactly if B", or similar things we are saying that A and B are basically the same statement. Formal notations for this is $A \leftrightarrow B$. We have often used $\equiv$ for this operator elsewhere[3], and the notation of Marcel (==) is

[3]which is an abuse, though others have used the symbol this way: the usual meaning of $A \equiv B$ is that $A \leftrightarrow B$ is a tautology (A and B are logically equivalent)

motivated by this alternative notation. "A iff B" is a learned abbreviation for "A if and only if B" which is used in mathematical English.

Proof Strategy: To deduce a goal of the form $A \leftrightarrow B$, deduce $A \rightarrow B$ and deduce $B \rightarrow A$. Since there are at least two strategies for deducing these implications, there are a number of ways to structure the proof.

One can use a posit of the form $A \leftrightarrow B$ in a number of ways. From posits $A \leftrightarrow B$ and A, we can deduce B; from posits $A \leftrightarrow B$ and B we can deduce A. More powerfully, if we have posits $A \leftrightarrow B$ and some complex $C[A]$, we can deduce $C[B]$ (simply replace occurrences of A with B) or symmetrically from posits $A \leftrightarrow B$ and $C[B]$ we can deduce $C[A]$[4].

The operation represented by $\leftrightarrow$ is called the *biconditional*.

We note without pursuing the details at this point that $A \not\leftrightarrow B$ (another commonly used notation is $A \oplus B$) is our notation for the "exclusive or": A or B is true but not both.

A common format for a theorem is to give a list of statements and assert that all of them are equivalent. A strategy for proving that statements $A_1, \ldots, A_n$ are equivalent is to show that $A_i \rightarrow A_{i+1 \bmod n}$ for each appropriate i (showing that each statement implies the next in a cycle). In a theorem of this type several linked cycles may be present.

We note that $(A \leftrightarrow B) \leftrightarrow C$ is equivalent to $A \leftrightarrow (B \leftrightarrow C)$ but *not* equivalent to $(A \leftrightarrow B) \wedge (B \leftrightarrow C)$ (there is an exercise about this later).[5]

1.6 Negation and Indirect Proof

It is common to say that the logical operation of negation (the formal notation is $\neg A$) means "not A". But "not A" is not necessarily an English sentence if A is an English sentence. A locution that works is "It is not the case that A", but we do not in fact usually say this in either everyday or mathematical English.

[4] As a matter of pedagogy, we prefer that students not use the substitution rule for biconditionals in homework proofs in the Proof part of the book. We will indicate specifically if we are allowing its use, or the use of specific kinds of substitution.

[5] But $A \equiv B \equiv C$ does mean "$A \leftrightarrow B$ is a tautology and $B \leftrightarrow C$ is a tautology", following the convention explained in the conjunction section.

"Not snow is white" is ungrammatical; "It is not the case that snow is white" is pedantic; "Snow isn't white" is what we say. If R is a relation symbol, we will often introduce a new relation symbol $\not R$ and let $x \not R y$ be synonymous with $\neg x\, R\, y$. The use of $\neq$ and $\notin$ should already be familiar to the reader.

We do not as a rule negate complex sentences in English. It is possible to say "It is not the case that both A and B are true" but this is only a formal possibility: what we would really say is "A is false or B is false". It is possible to say "It is not the case that either A or B is true" but this is also only a formal possibility: what we would really say is "A is false and B is false". The logical facts underlying these locutions are the identities

$$\neg(A \wedge B) \leftrightarrow (\neg A \vee \neg B)$$

and

$$\neg(A \vee B) \leftrightarrow (\neg A \wedge \neg B),$$

which are known as *de Morgan's laws*. It is pure common sense that we do not need to say "It is not the case that it is not the case that A", when we can so easily say A (the principle of double negation $\neg\neg A \leftrightarrow A$). $\neg(A \to B) \leftrightarrow A \wedge \neg B$ and $\neg(A \leftrightarrow B) \leftrightarrow (A \not\leftrightarrow B)$ might require a little thought. The former is best approached via the equivalence of $A \to B$ and $\neg A \vee B$ (which might itself require thought); the result about the negation of $A \to B$ then follows from de Morgan's laws and double negation. Do please note that we do not here authorize the use of these equivalences as proof strategies (without proof): they are mentioned here only as part of our discussion of the rhetoric of negation in mathematical English![6]

We present a brief example from algebra. To say $\neg(0 \leq x < 3)$ would be odd. We analyze this step by step. The chained relations hide a conjunction: $\neg(0 \leq x \wedge x < 3)$. De Morgan's law gives us $\neg 0 \leq x \vee \neg x < 3$. Negating the binary predicates rather than the atomic formulas gives us $0 > x \vee x \geq 3$, and a further obvious (non-logical) transformation gives us $x < 0 \vee x \geq 3$. Carrying out this kind of transformation reliably is expected of students in precalculus!

A statement of the form $A \wedge \neg A$ is called a *contradiction*. It is clear that such statements are always false. It is a logical truth that $A \vee \neg A$ is always true (this is called the law of *excluded middle*).

[6]This is a special case of our generally not allowing the use of the substitution rule for biconditionals in homework proofs in this part of the book.

We introduce the notation $\perp$ for a fixed false statement, which we may call "the absurd".

Proof Strategies:

1: To deduce a goal of the form $\neg A$, add A to your assumptions and deduce $\perp$, the absurd statement. Notice that we will certainly withdraw the assumption A and any posits deduced from it when this proof is done! We call this rule *negation introduction*.

2: From A and $\neg A$, deduce $\perp$. The only way to deduce the absurd is from a contradiction. We call this rule *contradiction*.

3: From $\neg\neg A$, deduce A. Otherwise, one can only use a negative hypothesis if the current goal is $\perp$: if we have a posit $\neg A$, use it by adopting A as a goal ("for the sake of a contradiction", so that $\perp$ can be deduced). We call this rule *double negation elimination*.

The first strategy above is *not* the notorious technique of "proof by contradiction": it is the direct strategy for proving a negative sentence. The strategy of proof by contradiction differs from all our other techniques in being applicable to sentences of any form: it can be viewed as the strategy of last resort.

Proof by Contradiction (reductio ad absurdum): To deduce any goal A at all, assume $\neg A$ and reason to $\perp$ (by reasoning to a contradiction). Notice that this is the same as a direct proof of the goal $\neg\neg A$. Our formal name for this rule is the classical *reductio ad absurdum*, since we have a rule above called "contradiction".

Principle of Double Negation: $\neg\neg P \leftrightarrow P$

Proof: Part 1 of the proof requires us to deduce P given the assumption $\neg\neg P$: this is given as a basic proof step above. Part 2 requires us to deduce $\neg\neg P$ given the assumption P: to do this, assume $\neg P$ and deduce $\perp$: but this is immediate as we have already assumed P. The proof is complete.

The derived rule "from P, deduce $\neg\neg P$", may be called *double negation introduction*.

In later parts of the book we will not usually mention $\bot$, so the strategy for proving $\neg A$ will generally be to deduce some contradiction $B \wedge \neg B$ from A (from which the further deduction of $\bot$ is immediate), and the strategy of proof by contradiction of A will be to deduce some contradiction $B \wedge \neg B$ from $\neg A$ (thus the name).

We prove that $P \to Q$ is equivalent to $\neg Q \to \neg P$. This will give our first extended example of the proof techniques we are advertising.

Contrapositives Theorem: $(P \to Q) \leftrightarrow (\neg Q \to \neg P)$

Proof: This breaks into two subgoals: Goal 1 is to prove $(P \to Q) \to (\neg Q \to \neg P)$ and Goal 2 is to prove $(\neg Q \to \neg P) \to (P \to Q)$.

We prove Goal 1: $(P \to Q) \to (\neg Q \to \neg P)$.

This goal is an implication, so we assume for the sake of argument that $P \to Q$: our new goal is $\neg Q \to \neg P$.

The new goal is also an implication, so we assume $\neg Q$ and have our latest goal as $\neg P$.

To deduce $\neg P$ we need to assume P and deduce $\perp$. We duly assume P. We have already assumed $P \to Q$, so modus ponens allows us to conclude Q. We have already assumed $\neg Q$, so we can conclude $\perp$, which is the goal, which allows us to complete the deduction of our latest goal $\neg P$, and so of the intermediate goal $\neg Q \to \neg P$ and so of Goal 1.

Goal 2 remains to be proved: $(\neg Q \to \neg P) \to (P \to Q)$. To prove this we need to assume $(\neg Q \to \neg P)$ and deduce an intermediate goal $P \to Q$. To deduce this goal, we need to assume P and deduce a second intermediate goal Q. To prove Q, we assume $\neg Q$ and take as our final intermediate goal $\perp$ (this is proof by contradiction). From $\neg Q$ and the earlier assumption $\neg Q \to \neg P$ we can conclude $\neg P$ by modus ponens. From the earlier assumption P and the recently proved $\neg P$ we conclude $\perp$, completing the deductions of all outstanding goals and the proof of the entire theorem.

We present the proof of the Contrapositives Theorem a second time in a more exhaustive style with all lines carefully labelled and references to rules used made explicit.

Contrapositives Theorem: $(P \to Q) \leftrightarrow (\neg Q \to \neg P)$

Proof: The statenment to be proved is a biconditional. This dictates a proof plan in two parts.

Part I: Assume (1): $P \to Q$

Each assumption gets a line number, because it is a posit and can be referenced in later applications of rules.

Goal: $\neg Q \to \neg P$

A goal does not get a line number, but it plays an important role in proof planning. In computer programming terms, it can be thought of as a comment.

Assume (2): $\neg Q$

Goal: $\neg P$

Assume (3): P

Goal: $\bot$

We have run out of ways to unpack our goals: we need to look for a way to use our posits. An opportunity presents itself!

(4): Q m.p. 1, 3 (m.p. abbreviates modus ponens)

(5): $\bot$, contradiction 2,4

(6:) $\neg P$ negation introduction 3-5

In earlier versions of our logic style manual, we tended to omit these closing lines, assuming that it is clear when goals have actually been met. We have learned that students prefer closure!

(7): $\neg Q \to \neg P$ deduction 2-6

(optional) (8): $(P \to Q) \to (\neg Q \to \neg P)$ deduction 1-7

We will present two ways of closing the entire argument, one using explicit references to the two implications making up the biconditional, and one which uses references to the two blocks.

Part II: Assume (9): $\neg Q \to \neg P$

The line number we use here is set of course after we see how long the first part is. It could have been 8, if we didn't use the last optional line. It could also of course have been written at the beginning as something like 1b. All that matter about line numbers is that different lines have different numbers, or at least that different lines we are actually entitled to reference have different numbers.

Goal: $P \to Q$

Assume (10): P

Goal: Q

Now we are in a pickle. The goal has no helpful structure and there is no obvious way to use the two posits in concert (we cannot use *modus tollens* because in fact we are proving a theorem intended to justify *modus tollens*.) When in doubt, use reductio ad absurdum!

Assume (11): $\neg Q$ for the sake of a contradiction.

Goal: $\perp$

(12): $\neg P$ m.p. 9,11

(13): $\perp$ contradiction 10,12

(14): Q reductio ad absurdum 11-13

(15): $P \to Q$ deduction 10-14

(optional)(16): $(\neg Q \to \neg P) \to (P \to Q)$ deduction 9-15

(17): $(P \to Q) \leftrightarrow (\neg Q \to \neg P)$ biconditional introduction 1-7, 9-15, **or** biconditional introduction 8, 16. Either style is acceptable; of course, if you use the first there is no reason to record line 8 or line 16.

Notice that we could replace the propositional letters P and Q with any statements A and B, however complex, and the proof above would still work: we have actually proved $(A \to B) \leftrightarrow (\neg B \to \neg A)$. This kind of generalization is the subject of a subsection below.

This justifies proof strategies we have already signalled above.

Proof Strategy: To prove a statement $A \to B$, we can aim instead for the equivalent $\neg B \to \neg A$: assume $\neg B$ and take $\neg A$ as our new goal. This is called "proving the contrapositive".

If we have posited both $A \to B$ and $\neg B$, then replacing the implication with the equivalent $\neg B \to \neg A$ and applying modus ponens allows us to conclude $\neg A$. The rule "From $A \to B$ and $\neg B$, conclude $\neg A$" is called *modus tollens*, and we have justified it.

We prove another theorem which justifies some additional proof strategies involving disjunction.

Theorem: $P \vee Q \leftrightarrow \neg P \to Q$

Corollary: $P \vee Q \leftrightarrow \neg Q \to P$. This follows from the theorem by equivalence of implications with their contrapositives and double negation.

Proof of Theorem: For part 1 of the proof, we assume $P \vee Q$ and deduce Goal 1: $\neg P \to Q$. The form of the posit suggests a proof by cases.

Case 1: We assume P. We prove the contrapositive of Goal 1: we assume $\neg Q$ and our goal is $\neg\neg P$. To prove $\neg\neg P$, we assume $\neg P$ and our goal is $\perp$, which is immediate as we have already posited P. This completes the proof of case 1.

Case 2: We assume Q. To prove the goal $\neg P \to Q$, we assume $\neg P$ and our new goal is Q. But we have already posited Q so we are done.

For part 2 of the proof, we assume $\neg P \to Q$ and deduce $P \vee Q$. We prove the goal by contradiction: we assume $\neg(P \vee Q)$ and take $\perp$ as our goal. We do this by proving P then proving $\neg P$. Our first goal is P, which we prove by contradiction: assume $\neg P$; by modus ponens Q follows, from which we can deduce $P \vee Q$, from which with our assumption $\neg(P \vee Q)$ we can deduce $\perp$, completing the proof of P by contradiction. Our second goal is $\neg P$: to prove this we assume P and take $\perp$ as our goal; from the assumption P we can deduce $P \vee Q$ from which with our assumption $\neg(P \vee Q)$ we can deduce $\perp$; this completes the proof of $\neg P$, which completes the proof by contradiction of $P \vee Q$.

Since the implications in both directions have been proved, the proof of the Theorem is complete.

Again, we present the theorem in a more exhaustive (exhausting?) format. This proof is actually rather different from the less formal proof given above; you could try formalizing the preceding proof as an exercise.

Theorem: $P \vee Q \leftrightarrow \neg P \rightarrow Q$

Proof: The statement is a biconditional, and gets the usual proof plan for a biconditional.

Part I: Assume (1): $P \vee Q$

Goal: $\neg P \rightarrow Q$

Assume (2): $\neg P$

Goal: Q

Assume (3): Assume $\neg Q$ for the sake of a contradiction

Goal: $\perp$

We start a proof by cases using line 1.

Case I: assume (4a): P

Goal: $\perp$

(5a): $\perp$ contradiction, 2,4a

Case 2: assume (4b): Q

Goal: $\perp$

(5b): $\perp$ contradiction 3,4b

(6): $\perp$ proof by cases, 1, 4a-5a, 4b-5b (proof by cases has the most complex line justifications of any of our rules!)

(7): Q reductio ad absurdum 3-6

(8): $\neg P \rightarrow Q$ deduction 2-7

Part II: Assume (9): $\neg P \to Q$

Goal: $P \vee Q$

We start by throwing up our hands in despair and trying reductio ad absurdum!

Assume (10): $\neg(P \vee Q)$ for the sake of a contradiction.

Goal: $\perp$

We introduce a new goal $\neg P$ with the idea that if we can prove it, we can apply modus ponens with line 9 to prove Q.

Goal: $\neg P$

Assume (11): P

Goal: $\perp$

(12): $P \vee Q$ addition, line 11

(13): $\perp$ contradiction, 10, 12

(14) $\neg P$ negation introduction 11-13

(15): Q m.p. 9,14

(16): $P \vee Q$ addition 15

(17): $\perp$ contradiction 10, 16

(18): $P \vee Q$ reductio ad absurdum 10-17

(19): $(P \vee Q) \leftrightarrow (\neg P \to Q)$ biconditional introduction 1-8, 9-18.

This is a hard proof because the only rules for disjunction we are allowed to use are disjunction and proof by cases. Even excluded middle is proved using this theorem, and so cannot be assumed here.

We prove the symmetric form $(A \vee B) \leftrightarrow (\neg B \rightarrow A)$.

Theorem: $(A \vee B) \leftrightarrow (\neg B \rightarrow A)$.

Part I: Assume (1): $A \vee B$

Goal: $\neg B \rightarrow A$

Assume(2): $\neg B$

Goal: B

(3): $(A \vee B) \leftrightarrow (\neg A \rightarrow B)$. Previous theorem.

(4): $\neg A \rightarrow B$ bimp (biconditional m.p.) 1,3

(5): $(\neg A \rightarrow B) \leftrightarrow (\neg B \rightarrow \neg\neg A)$ contrapositives theorem, replacing P with $\neg A$, Q with B.

(6): $\neg B \rightarrow \neg\neg A$ bimp 4,5

(7): $\neg\neg A$ m.p. 2,6

(8): A dne (double negation elimination) 7

(9): $\neg B \rightarrow A$ deduction 2-8

Part II: Assume (10): $\neg B \rightarrow A$

Goal: $A \vee B$

Goal: $\neg A \rightarrow B$ (so we can apply the previous theorem)

Assume(11): $\neg A$

Goal: B

(12): $\neg A \rightarrow \neg\neg B$ contrapositives theorem 10 (this time just using the theorem as a one step rule, "deduce $\neg D \rightarrow \neg C$ from $C \rightarrow D$".)

(13:) $\neg\neg B$ m.p. 11, 12

(14): B dne 13

(15): $\neg A \rightarrow B$ deduction 11-14

(16): $A \vee B$ (the previous theorem, used as a one step rule, "deduce $C \vee D$ from $\neg C \rightarrow D$")

(17:) $(A \vee B) \leftrightarrow (\neg B \rightarrow A)$ biconditional introduction, 1-9, 10-16.

Note that in Part I of this proof we exhibit a very conservative way of appealing to a theorem already proved, and in Part II we illustrate a more liberal way of doing this.

The Theorem directly justifies the more general proof strategies for disjunction involving negative hypotheses given above.

Proof Strategy: To deduce the goal $A \vee B$, assume $\neg A$ and deduce B: this is valid because it is a proof of the equivalent implication $\neg A \rightarrow B$. Alternatively, assume $\neg B$ and deduce A: this is a proof of the equivalent $\neg B \rightarrow A$. These rules are called *disjunction introduction* (or *alternative elimination*)

If we have posits $A \vee B$ and $\neg A$, we can draw the conclusion B, by converting $A \vee B$ to the equivalent $\neg A \rightarrow B$ and applying modus ponens.

Symmetrically, if we have posits $A \vee B$ and $\neg B$, we can deduce A.

The latter two rules are called *disjunctive syllogism*. We also view two other variants as instances of disjunctive syllogism: "given $A \vee \neg B$ and B, deduce A", and "given $\neg A \vee B$ and A, deduce B". Of course these follow from the other forms and applications of double negation introduction.

A classic theorem which we should not neglect, often used as the basis for proofs by cases, is

Theorem (excluded middle): $A \vee \neg A$

Proof: To prove this by the alternative elimination strategy, assume $\neg\neg A$ and show that A follows. But it does follow, immediately, by double negation elimination.

A perhaps shocking result is that anything at all follows from $\perp$, and so from any contradiction.

Theorem: $\perp \rightarrow B$

Proof: Assume $\perp$, and our goal becomes B. We prove B by contradiction, that is, assume $\neg B$ and take $\perp$ as our new goal. The new goal is already met by our initial assumption, so the proof is complete.[7]

[7] In a constructive logic, where double negation elimination is not allowed as a rule, deduction of any B from $\perp$ would be a primitive rule, called *absurdity elimination*.

Theorem: $A \wedge \neg A \rightarrow B$

Proof: Assume $A \wedge \neg A$, and take B as our new goal. From $A \wedge \neg A$, we deduce A and we deduce $\neg A$, and from these we deduce $\bot$. $\bot \rightarrow B$ is true by the previous theorem, and B follows by *modus ponens*.

The operation represented by $\neg$ is called *negation*.

1.7 Generality and the Rule of Substitution

A propositional letter P reveals nothing about the structure of the statement it denotes. This means that any argument that shows that certain hypotheses (possibly involving P) imply a certain conclusion A (possibly involving P) will remain valid if all occurrences of the propositional letter P in the entire context are replaced with any fixed statement B (which may be logically complex).

Denote the result of replacing P with B in A by $A[B/P]$. Extend this notation to sets Γ: $\Gamma[B/P] = \{A[B/P] \mid A \in \Gamma\}$.

The rule of substitution for propositional logic can then be stated as

If we can deduce A from a set of assumptions Γ, P is a propositional letter and B is any proposition (possibly complex), then we can deduce $A[B/P]$ from the assumptions $\Gamma[B/P]$.

Using the substitution notation, the strongest rules for the biconditional can be stated as

"from $A \leftrightarrow B$ and $C[B/P]$, deduce $C[A/P]$."

"from $A \leftrightarrow B$ and $C[A/P]$, deduce $C[B/P]$."

1.8 Note on Order of Operations

The statements "A and either B or C" and "Either A and B, or C" (which can formally be written $A \wedge (B \vee C)$ and $(A \wedge B) \vee C)$ do not have the same meaning. Making such grouping distinctions in English is awkward; in our notation we have the advantage of the mathematical device of parentheses.

To avoid having to write all parentheses in order to make the meaning of a statement clear, we stipulate that just as multiplication is carried out before addition when parentheses do not direct us to do otherwise, we carry out $\neg$ first, then $\wedge$, then $\vee$, then $\rightarrow$, then $\leftrightarrow$ or $\not\leftrightarrow$. When a list of operations

at the same level are presented, we group to the right: $P \to Q \to R$ means $P \to (Q \to R)$. In fact, this only makes a difference for $\to$, as all the other basic operations are associative (including $\leftrightarrow$ and $\not\leftrightarrow$; check it out!).[8]

There is a temptation to allow $A \leftrightarrow B \leftrightarrow C$ to mean $(A \leftrightarrow B) \wedge (B \leftrightarrow C)$ by forbidding the omission of parentheses in expressions $A \leftrightarrow (B \leftrightarrow C)$ and $(A \leftrightarrow B) \leftrightarrow C$. We resist this temptation.

1.9 Quantifiers

In this section, we go beyond propositional logic to what is variously called first-order logic, predicate logic, or the logic of quantifiers. In any event, as in the propositional logic section, we are not talking about a formal system, though we will introduce some formal notations: we are talking about kinds of statement which appear in informal mathematical argument in natural language, and introducing formal notations to represent such statements in aid of precision, and we hope in aid of clarity.

We denote an arbitrary complex statement, presumably involving the variable x, by the notation $A[x]$. We do not write $A(x)$ because this is our notation for a unary predicate sentence in which A stands for some definite unary predicate: a sentence of the form $A(x)$ has the exact form of a predicate being asserted of x while a sentence of the form $A[x]$ could be any sentence that presumably mentions x (so $x = x$ is of the form $A[x]$ but not of the form $A(x)$; a sentence like $\mathtt{Nat}(x)$ (meaning "x is a natural number") would be an example of the first form. A related notation is $A[t/x]$, the result of replacing the variable x in the proposition A with the term t (which may be a complex name rather than simply a variable). If we denote a formula $\mathcal{A}$ by the notation $A[x]$ then for any term t we use the notation $A[t]$ to represent $\mathcal{A}[t/x]$.[9]

The two kinds of statement we consider can be written "for all x, $A[x]$" (formulas with a *universal quantifier*) and "for some x, $A[x]$", which is also often written "there exists x such that $A[x]$" (which is why such formulas

[8]The theorem proving software Marcel treats conjunction and disjunction as grouping to the left by default for reasons which will become clear to the observant when they actually carry out some proofs. But this is a matter of convenience.

[9]It should be noted that this is a subtle distinction I am drawing which is not universally made (the exact notation here is specific to these notes); it is quite common to write $P(x)$ for what I denote here as $P[x]$, and I have been known to write parentheses by mistake when teaching from this text.

are said to have an *existential quantifier*). This language, although it is acceptable mathematical English, is already semi-formalized.

Formulas (or sentences) with universal and existential quantifiers can appear in a variety of forms. The statement "All men are mortal" can be analyzed as "for all x, if x is a man then x is mortal", and the statement "Some men are immortal" can be analyzed as "for some x, x is a man and x is immortal".

The formal notation for "for all x, $A[x]$" is $(\forall x.A[x])$ and for "for some x, $A[x]$" is $(\exists x.A[x])$. The parentheses in these notations are for us mandatory: this may seem eccentric but we think there are good reasons for it.[10]

Iteration of the same quantifier can be abbreviated. We write $(\forall xy.A[x,y])$ instead of $(\forall x.(\forall y.A[x,y]))$, and similarly $(\exists xy.A[x,y])$ instead of $(\exists x.(\exists y.A[x,y]))$, and notations like $(\forall xyz.A[x,y,z])$ are defined similarly.

Quantifiers are sometimes (very often, in practice), *restricted* to some domain. Quantifiers restricted to a set have special notation: $(\forall x \in S.A[x])$ can be read "for all x in S, $A[x]$" and is equivalent to $(\forall x.x \in S \to A[x])$, while $(\exists x \in S.A[x])$ can be read "for some x in S, $A[x]$" and is equivalent to $(\exists x.x \in S \wedge A[x])$. The same quantifier restricted to the same set can be iterated, as in $(\forall xy \in S.A[x,y])$, meaning $(\forall x \in S.(\forall y \in S.A[x,y]))$. We leave the expansion of this with implications to the imagination. Restriction can also use other binary predicates: $(\forall x\, R\, y.A[x]\}$ abbreviates $(\forall x.x\, R\, y \to A[x])$, for example.

Further, restriction of a quantifier to a particular sort of object is not always explicitly indicated in the notation. If we know from the context that a variable n ranges over natural numbers, we can write $(\forall n.A[n])$ instead of $(\forall n \in \mathcal{N}.A[n])$, for example. In the chapter on typed theory of sets, all variables will be equipped with an implicit type in this way.

We do not as a rule negate quantified sentences (or formulas) in natural language. Instead of saying "It is not the case that for all x, $A[x]$", we would say "For some x, $\neg A[x]$". Instead of saying "It is not the case that for some x, $A[x]$", we could say "For all x, $\neg A[x]$" (though English provides us with the construction "For no x, $A[x]$" for this case). "No men are mortal" means "For all x, if x is a man then x is not mortal". The logical transformations which can be carried out on negated quantified sentences are analogous to

[10]Nowadays (2021) I tend to write $(\forall x : P[x])$ and $(\exists x : P[x])$, and this form may be seen in more recent edits of the text.

de Morgan's laws, and can be written formally

$$\neg(\forall x.A[x]) \leftrightarrow (\exists x.\neg A[x])$$

and

$$\neg(\exists x.A[x]) \leftrightarrow (\forall x.\neg A[x]).$$

Note that we are not licensing use of these equivalences as proof strategies before they are proved: as above with de Morgan's laws, these are introduced here to make a point about the rhetoric of mathematical English.

Here is a good place to say something formally about the distinction between the more general "formula" and the technical sense of "sentence" (I would really much rather say "sentence" for both, following the grammatical rather than the mathematical path). Any "sentence" in the grammatical sense of mathematical language is called a formula; the actual "sentences" in the mathematical sense are those in which a variable x only occurs in a context such as $(\forall x.A[x])$, $(\exists x.A[x])$ or even $\{x \mid A[x]\}$ or $\int_2^3 x^2\,dx$ (to get even more familiar) in which it is a dummy variable. The technical way of saying this is that a sentence is a formula in which all occurrences of variables are *bound*.

1.9.1 Variation: Hilbert symbols and definite descriptions

A variant approach to quantification is supported by the use of the *Hilbert symbol* $(\epsilon x : A[x])$, which may be read "An x such that $A[x]$ (if there is one)." This symbol stands for an arbitrarily selected x such that $A[x]$, if there is one, and otherwise for a default object. The thing to notice here is that on the intended semantics for the Hilbert symbol, $(\exists x : A[x])$ is equivalent to $A[(\epsilon x : A[x])]$. Thus the existential quantifier can be defined in terms of the Hilbert symbol, and indeed $(\forall x : A[x])$ can also be expressed as $A[(\epsilon x : \neg A[x])]$. An attempt to expand any expression with nested quantifiers should reveal why these are not really good practical definitions, but this notion has extensive theoretical uses.

A *definite description* $(\theta x : A[x])$ (read "The x such that $A[x]$)" stands for the unique x such that $(\theta x : A[x])$ (if there is such an x): $(\theta x : A[x])$ could be defined in terms of the Hilbert symbol as $(\epsilon x : (\forall y : A[y] \leftrightarrow y = x))$.

We postulate additionally (if we are using these symbols: they are not an official part of our logic except when we explicitly say so) that

$$(\forall x : A[x] \leftrightarrow B[x]) \rightarrow (\epsilon x : A[x]) = (\epsilon x : B[x]))$$

and similarly

$$(\forall x : A[x] \leftrightarrow B[x]) \rightarrow (\theta x : A[x]) = (\theta x : B[x])) :$$

the object such that $A[x]$ and the object such that $B[x]$ which are arbitarily chosen are the same object if the two sentences are true of the same things.[11] This has the side effect that for any $A[x]$ which is not true for any x, $(\epsilon x : A[x])$ is the same default object (and we stipulate that this is the same as $(\theta x : B[x])$ for all $B[x]$ which hold for no x or for more than one x). We will temporarily refer to this default object as δ: a good choice for δ is the empty set $\emptyset$.

1.10 Proving Quantified Statements and Using Quantified Hypotheses

To prove the goal "for all x, $A[x]$", introduce a new name a (not used before anywhere in the context): the new goal is to prove $A[a]$. Informally, if we can prove $A[a]$ without making any special assumptions about a, we have shown that $A[x]$ is true no matter what value the variable x takes on. The new name a is used only in this proof (its role is rather like that of an assumption in the proof of an implication). This rule is called "universal generalization".

It is permissible and sometimes useful to open the indented block in a proof of "for all x, $A[x]$" using the new goal $A[a]$ with a numbered line $a = a$: it is handy to have a first line to reference the start of the block in a uniform way.

To prove the goal "for some x, $A[x]$", find a specific name t (which may be complex) and prove $A[t]$. Notice here there may be all kinds of contextual knowledge about t and in fact that is expected. It's possible that several different such substitutions may be made in the course of a proof (in different cases a different witness may work to prove the existential statement). This rule is called "existential instantiation".

[11] This serves to preserve the validity of the rule that logically equivalent formulas can replace one another freely.

If you have posited "for all x, $A[x]$", then you may further posit $A[t]$ for any name t, possibly complex. You may want to make several such substitutions in the course of a proof. This rule is called "universal instantiation".

Using an existential statement is a bit trickier. If we have posited "for some x, $A[x]$", and we are aiming at a goal G, we may introduce a name w not mentioned anywhere in the context (and in particular not in G) and further posit $A[w]$: if G follows with the additional posit, it follows without it as well. What we are doing here is introducing a name for a witness to the existential hypothesis. Notice that this name is locally defined; it is not needed after the conclusion G is proved. This rule is called "existential generalization" or "witness introduction".

We present an example.

Theorem: $(\exists x : A[x]) \wedge (\forall x : A[x] \rightarrow B[x]) \rightarrow (\exists x : B[x])$

Proof:

Assume (1): $(\exists x : A[x]) \wedge (\forall x : A[x] \rightarrow B[x])$
Goal: $(\exists : B[x])$
(2): $(\exists x : A[x])$ simplification 1
(3): $(\forall x : A[x] \rightarrow B[x])$ simplification 1
(4): $A[w]$ introduce a witness to line 2
Goal: $(\exists x : B[x])$ as is usual with EG, this is the goal you already have
(5): $A[w] \rightarrow B[w]$ universal instantiation, $x := w$, line 3
(6): $B[w]$ mp 4,5
(7): $(\exists x : B[x])$ existential instantiation $x := w$ line 6
(8): $(\exists x : B[x])$ existential generalization 4-7
(9): $(\exists x : A[x]) \wedge (\forall x : A[x] \rightarrow B[x]) \rightarrow (\exists x : B[x])$ deduction 1-8

1.10.1 Reasoning with the Hilbert symbol

One of the merits of the Hilbert symbol is that it might help to demystify the patterns of reasoning above, if one did happen to be mystified.

The sole relevant reasoning rule for the Hilbert symbol is "From $A[t]$, deduce $A[(\epsilon x : A[x])$", where t is any expression and x is a variable not appearing in $A[t]$.

Suppose that we can prove $A[a]$ where a is a brand-new constant used nowhere before. Then we can define a retrospectively as $(\epsilon x : \neg A[x])$ and we find that $A[a]$ literally is $(\forall x : A[x])$ if we use the definition of the universal quantifier given above.

If we assume $(\forall x : A[x])$, then we have assumed $A[(\epsilon x : \neg A[x])]$. We argue by contradiction that we can conclude $A[t]$: if we on the contrary suppose $\neg A[t]$, we can deduce $\neg A[(\epsilon x : \neg A[x])]$ using the basic rule for the Hilbert symbol, which yields a contradiction.

Suppose that we can prove $A[t]$: then by the basic rule for reasoning with the Hilbert symbol we have $A[(\epsilon x : A[x])]$, thus $(\exists x : A[x])$.

Suppose that we assume $(\exists x : A[x])$. If we introduce a new symbol w and from $A[w]$ deduce a conclusion G not mentioning w, we can retrospectively define w as $(\epsilon x : A[x])$ and we suddenly realize that we have deduced G from $A[w]$ which simply is the hypothesis $(\exists x : A[x])$.

1.11 Equality and Uniqueness

For any term t, $t = t$ is an axiom which we may freely assert.

If we have posited $a = b$ and $A[a/x]$, we can further posit $A[b/x]$.

We recall from above that if we include the Hilbert symbol in our logic, we add the rule "if we have posited $(\forall x : A[x] \leftrightarrow B[x])$, we may further posit $(\epsilon x : A[x]) = (\epsilon x : B[x])$."

These are an adequate set of logical rules for equality.

To show that there is exactly one object x such that $A[x]$ (this is often written $(\exists! x.A[x])$), one needs to show two things: first, show $(\exists x.A[x])$ (there is at least one x). Then show that from the additional assumptions $A[a]$ and $A[b]$, where a and b are new variables not found elsewhere in the context, that we can prove $a = b$ (there is at most one x).[12]

Proofs of uniqueness are often given in the form "Assume that $A[a]$, $A[b]$, and $a \neq b$: deduce a contradiction". This is equivalent to the proof strategy just given but the assumption $a \neq b$ is often in practice never used (one simply proves $a = b$) and so seems to be an unnecessary complication.

The rules of symmetry and transitivity of equality are consequences of the rules given above.

We demonstrate the validity of the rules

[12]Note that under the hypothesis $(\exists! x.A[x])$, the definite description notation $(\theta x : A[x])$ has the intended meaning. A logically rather opaque way of expressing the same condition is "$(\theta x : A[x])$ exists", though on our convention the definite description denotes a definite object δ even if it does not have the intended meaning.

$$\frac{a = b}{b = a}$$

and

$$\frac{\begin{array}{c} a = b \\ b = c \end{array}}{a = c}$$

We demonstrate the validity of the first rule.

(1): $a = b$ premise

(2): $a = a$ reflexivity of equality

(3): $b = a$ substitution into 2 using $x = a$ as the formula A: $A[a/x]$ is $a = a$ and $A[b/x]$ is $b = a$. Notice that this way of describing substitution allows us to deal with situations where we do not want to replace all a's with b's.

We demonstrate the validity of the second rule.

(1): $a = b$ premise

(2): $b = c$ premise

(3): $b = a$ symmetry of equality (the rule just proved)

(4): $[a] = c$ substitution into 2 using 3 (using the bracket to highlight where the substitution happened).

1.12 Dummy Variables and Substitution

The rules of the previous section make essential use of substitution. If we write the formula $A[x]$ of the previous section in the form $\mathcal{A}$, recall that the variants $A[a]$ and $A[t]$ mean $\mathcal{A}[a/x]$ and $\mathcal{A}[t/x]$: understanding these notations requires an understanding of substitution.

And there is something nontrivial to understand. Consider the sentence $(\exists x.x = a)$ (this is a sentence if a is a constant *name* rather than a variable). This is true for any a, so we might want to assert the not very profound

theorem $(\forall y.(\exists x.x = y))$. Because this is a universal statement, we can drop the universal quantifier and replace y with anything to get another true statement: with c to get $(\exists x.x = c)$); with z to get $(\exists x.x = z)$). But if we naively replace y with x we get $(\exists x.x = x)$, which does not say what we want to say: we want to say that there is something which is equal to x, and instead we have said that there is something which is equal to *itself*.

The problem is that the x in $(\exists x.x = y)$ does not refer to any particular object (even if the variable x does refer to something in a larger context). x in this sentence is a "dummy variable". Since it is a dummy it can itself be replaced with any other variable: $(\exists w.w = y)$ means the same thing as $(\exists x.x = y)$, and replacing y with x in the former formula gives $(\exists w.w = x)$ which has the intended meaning.

Renaming dummy variables as needed to avoid collisions avoids these problems. We give a recursive definition of substitution which supports this idea. $T[t/x]$ is defined for T any term or formula, t any term, and x any variable. The only kind of term (noun phrase) that we have so far is variables: $y[t/x]$ is y if $y \neq x$ and t otherwise; $P(u)[t/x]$ is $P(u[t/x])$; $(u\,R\,v)[t/x]$ is $u[t/x]\,R\,v[t/x]$. So far we have defined substitution in such a way that it is simply replacement of the variable x by the term t. Where A is a formula which might contain x, $(\forall y.A)[t/x]$ is defined as $(\forall z.A[z/y][t/x])$, where z is the typographically first variable not occurring in $(\forall y.A)$, t or x. $(\exists y.A)[t/x]$ is defined as $(\exists z.A[z/y][t/x])$, where z is the typographically first variable not occurring in $(\exists y.A)$, t, or x. This applies to all constructions with bound variables, including term constructions: for example, once we introduce set notation, $\{y \mid A\}[t/x]$ will be defined as $\{z \mid A[z/y][t/x]\}$), where z is the typographically first variable not occurring in $\{y \mid A\}$, t, or x. The use of "typographically first" here is purely for precision: in fact our convention is that (for example) $(\forall x.A)$ is basically the same statement as $(\forall y.A[y/x])$ for any variable y not occurring in A (where our careful definition of substitution is used) so it does not matter which variable is used as long as the variable is new in the context.

It is worth noting that the same precautions need to be taken in carefully defining the notion of substitution for a propositional letter involved in the rule of substitution.

We also note that the actual substitutions involved in expanding out any nontrivial reasoning with Hilbert symbols would create very large expressions

indeed![13]

1.13 Are we doing formal logic yet?

One might think we are already doing formal logic. But from the strictest standpoint we are not. We have introduced formal notations extending our working mathematical language, but we are not yet considering terms, formulas and proofs *themselves* as mathematical objects and subjecting them to analysis (perhaps we are threatening to do so in the immediately preceding subsection). We will develop the tools we need to define terms and formulas as formal mathematical objects (actually, the tools we need to formally develop any mathematical object whatever) in the next section, and return to true formalization of logic (as opposed to development of formal notation) in the Logic chapter.

We have not given many examples: our feeling is that this material is so abstract that the best way to approach it is to use it when one has some content to reason about, which will happen in the next chapter. Reference back to our discussion of proof strategy here from actual proofs ahead of us is encouraged.[14]

[13] Note to self: add some exercises expanding on this point

[14] and lab work with Marcel will also give exanples of abstract reasoning with quantifiers.

1.14 Exercises

Prove the following statements using the proof strategies above. Use only the highlighted proof strategies (not, for example, de Morgan's laws or the rules for negating quantifiers, or the use of biconditional theorems to make substitutions). You may use proof of an implication by proving the contrapositive, modus tollens and the generalized rules for proving disjunctions.

1. Prove the equivalence

$$A \to (B \to C) \leftrightarrow (A \wedge B) \to C$$

2. Prove

$$((P \to Q) \wedge (Q \to R)) \to (P \to R)$$

   This should be very straightforward.

3. Prove

$$(A \wedge B) \leftrightarrow (B \wedge A)$$

   This should be very straightforward, and very annoying.

4. Prove the equivalence

$$\neg(A \to B) \leftrightarrow (A \wedge \neg B)$$

   This will be harder: remember, when you can't see anything else to do, use reductio.

5. Prove each of the following:

   (a)
$$\neg(\forall x.P[x]) \leftrightarrow (\exists x.\neg P[x])$$

   (b)
$$\neg(\exists x.P[x]) \leftrightarrow (\forall x.\neg P[x])$$

6. Prove

$$((\exists x.P[x]) \wedge (\forall uv.P[u] \to Q[v])) \to (\forall z.Q[z])$$

7. Prove de Morgan's laws (both of them).

8. Verify that

$$((A \vee B) \to C) \leftrightarrow ((A \to C) \wedge (B \to C))$$

is a theorem.

9. Verify the rule of *destructive dilemma*:

$$\frac{\begin{array}{c} P \to Q \\ R \to S \\ \neg Q \vee \neg S \end{array}}{\neg P \vee \neg R}$$

An example of verification of a related rule appears as an example in the manual of logical style at the end of the text.

10. Justify the rules for the existential quantifier using the rules for the universal quantifier and the equivalence of $(\exists x.P[x])$ and $\neg(\forall x.\neg P[x])$.

11. Construct truth tables for $A \leftrightarrow (B \leftrightarrow C)$, $(A \leftrightarrow B) \leftrightarrow C$, and

$$(A \leftrightarrow B) \wedge (B \leftrightarrow C).$$

Notice that the first two are the same and the third (which one might offhand think is what the first says) is quite different. Can you determine a succinct way of explaining what

$$A_1 \leftrightarrow A_2 \leftrightarrow \ldots \leftrightarrow A_n$$

says?

12. (added 2/4/2021) Give a paper proof of $(\exists x : (\forall y : P[y] \to P[x]))$. I'll try to provide notes comparing this proof with the Marcel proof we did in class. Then prove $(\exists x : (\forall y : P[x] \to P[y]))$. I might have you do this one in a Marcel lab.

This one might be really difficult.

13. (added 2/9/2021) Verify the validity of the argument

$$\frac{\begin{array}{c}(\forall x : \neg G[x] \vee \neg H[x]) \\ (\forall x : (J[x] \to F[x]) \to H[x])\end{array}}{\neg(\exists x : F[x] \wedge G[x])}$$

Hint: You will want to assume $(\exists x : F[x] \wedge G[x])$ and reason to a contradiction.

We give some solutions.

1: Our goal is

$$A \to (B \to C) \leftrightarrow (A \wedge B) \to C.$$

Goal 1:

$$A \to (B \to C) \to (A \wedge B) \to C$$

.

Argument for Goal 1: Assume $A \to (B \to C)$. Our new goal is $(A \wedge B) \to C$. To prove this implication we further assume $A \wedge B$, and our new goal is C. Since we have posited $A \wedge B$, we may deduce both A and B separately. Since we have posited A and $A \to (B \to C)$ we may deduce $B \to C$ by modus ponens. Since we have posited B and $B \to C$, we may deduce C, which is our goal, completing the proof of Goal 1.

Goal 2:

$$(A \wedge B) \to C \to A \to (B \to C)$$

Argument for Goal 2: Assume $(A \wedge B) \to C$. Our new goal is $A \to (B \to C)$. To deduce this implication, we assume A and our new goal is $B \to C$. To deduce this implication, we assume B and our new goal is C. Since we have posited both A and B we may deduce $A \wedge B$. Since we have posited $A \wedge B$ and $(A \wedge B) \to C$, we may deduce C by modus ponens, which is our goal, completing the proof of Goal 2.

Conclusion: Since the implications in both directions have been proved, the biconditional main goal has been proved.

5a: Our goal is

$$\neg(\forall x.P[x]) \leftrightarrow (\exists x.\neg P[x]).$$

Since this is a biconditional, the proof involves proving two subgoals.

Goal 1:

$$\neg(\forall x.P[x]) \rightarrow (\exists x.\neg P[x])$$

Argument for Goal 1: Assume $\neg(\forall x.P[x])$. Our new goal is $(\exists x.\neg P[x])$. We would like to prove this by exhibiting a witness, but we have no information about any specific objects, so our only hope is to start a proof by contradiction. We assume $\neg(\exists x.\neg P[x])$ and our new goal is $\perp$. We note that deducing $(\forall x.P[x])$ as a goal would allow us to deduce $\perp$ (this is one of the main ways to use a negative hypothesis). To prove this goal, introduce an arbitrary object a and our new goal is $P[a]$. Since there is no other evident way to proceed, we start a new proof by contradiction: assume $\neg P[a]$ and our new goal is $\perp$. Since we have posited $\neg P[a]$, we may deduce $(\exists x.\neg P[x])$. This allows us to deduce $\perp$, since we have already posited the negation of this statement. This supplies what is needed for each goal in turn back to Goal 1, which is thus proved.

Goal 2:

$$(\exists x.\neg P[x]) \rightarrow \neg(\forall x.P[x])$$

Argument for Goal 2: We assume $(\exists x.\neg P[x])$. Our new goal is $\neg(\forall x.P[x])$. To deduce this goal, we assume $(\forall x.P[x])$ and our new goal is $\perp$. Our existential hypothesis $(\exists x.\neg P[x])$ allows us to introduce a new object a such that $\neg P[a]$ holds. But our universal hypothesis $(\forall x.P[x])$ allows us to deduce $P[a]$ as well, so we can deduce $\perp$, completing the proof of Goal 2.

Conclusion: Since both implications involved in the biconditional main goal have been proved, we have proved the main goal.

We present this again in the more explicit format with line numbers and justifications.

Prove:

$$\neg(\forall x.P[x]) \leftrightarrow (\exists x.\neg P[x]).$$

Part I:

Assume (1): $\neg(\forall x.P[x])$

Goal: $(\exists x.\neg P[x])$

Assume (2): $\neg(\exists x.\neg P[x])$ for the sake of a contradiction

Goal: $\perp$

Goal: $(\forall x.P[x])$ (which will give a contradiction with 1)

Let a be arbitrary.

Goal: $P(a)$

Assume (3): $\neg P[a]$ for the sake of a contradiction

Goal: $\perp$

(4): $(\exists x.\neg P[x])$ EI, 3

(5): $\perp$ contradiction 2,4

(6) $P(a)$ RAA 3-5

(7): $(\forall x.P[x])$ UG 3-6

(8): $\perp$ contradiction 1,7

(9): $(\exists x.\neg P[x])$ RAA 2-8

Part II:

Assume (1b): $(\exists x.\neg P[x])$

Goal: $\neg(\forall x.P[x])$

Assume (2b) : $(\forall x.P[x])$, for the sake of a contradiction.

Goal: $\perp$

(3b): $\neg P(w)$ introduce witness to hypothesis 1b

(4b): $P(w)$ UI line 2b $x := w$

(5b): $\perp$ contradiction, 3b,4b

(6b): $\perp$ EG 3b-5b

(7b): $\neg(\forall x.P[x])$ neg intro 2b-6b

the result to be proved: follows by biconditional introduction, 1-9, 1b-7b.

12a: This proof presents a significant technical challenge which we illustrate. Thw difficulty is that there isn't any specific object we have introduced in the environment to use as a candidate to prove the statement by EI. Watch and marvel at the solution.

Goal: $(\exists x : (\forall y : P[y] \to P[x]))$

Goal: $(\forall z : (\exists x : (\forall y : P[y] \to P[x])))$ You will see why we introduce the extra quantifier, and how we get rid of it.

1: $a = a$ (let a be arbitary, but provide a line number).

Goal: $(\exists x : (\forall y : P[y] \to P[x]))$

2: $(\exists x : P[x]) \vee \neg(\exists x : P[x])$ excluded middle
we proceed with a proof by cases on line 2

Case 1 (3a): $(\exists x : P[x])$

(4a): $P[w]$ introduce a witness to line 3a

Goal: $(\forall y : P[y] \to P[w])$

5a: $b = b$ let b be arbitary

Goal: $P[b] \to P[w]$
(it wont let me indent this block)

Assume (6a): $P[b]$

Goal: $P[w]$

7a: $P[w]$ copied from line 4a
(end block)

8a: $P[b] \to P[w]$ deduction 6a-7a

9a: $(\forall y : P[y] \to P[w])$ UG 5a-8a

10a: $(\exists x : (\forall y : P[y] \to P[x]))$ EI 9a x:=w

11a: $(\exists x : (\forall y : P[y] \to P[x]))$ EG 4a-10a

Case 2: (3b): $\neg(\exists x : P[x])$

Goal: $(\forall y : P[y] \rightarrow P[a])$

4b $b = b$ let b be arbitrary

Goal: $P[b] \rightarrow P[a]$

Assume(5b): $P[b]$

6b: $(\exists x : P[x])$ EI 5b x:=b

7b: $\perp$ contradiction 3b,6b

8b: $\perp \rightarrow P[a]$ theorem proved in the text, false implies anything

9b: $P[a]$ mp 7b,8b

10b $P[b] \rightarrow P[a]$ deduction 5b-9b

11b $(\forall y : P[y] \rightarrow P[a])$ UG 4b-10b

12b: $(\exists x : (\forall y : P[y] \rightarrow P[x]))$ EI 11b x:=a

13: $(\exists x : (\forall y : P[y] \rightarrow P[x]))$ proof by cases, 2, 3a-11a, 3b-12b (I should have had 2a, 2b, but life is too short to change all the numbers)

14: $(\forall z : (\exists x : (\forall y : P[y] \rightarrow P[x])))$ UG 1-13

15: $(\exists x : (\forall y : P[y] \rightarrow P[x]))$ UI 14 (no need to say what z is assigned since no occurrence of z needs to be adjusted)

Isn't that weird? I could add a rule which allows introduction of an arbitrary object, but...I dont need one! This also provides extensive examples of the idea of adding a trivial line mentioning the arbitrary object at the beginning of a UG block.

Chapter 2

Typed theory of sets

In this chapter we introduce a theory of sets, but not the usual one quite yet. We choose to introduce a typed theory of sets, which might carelessly be attributed to Russell, though historically this is not quite correct.

2.1 Types in General

Mathematical objects come in sorts or kinds (the usual word is "type"). We seldom make any statement about all mathematical objects whatsoever: we are more likely to be talking about all natural numbers, or all real numbers, or all elements of a certain vector space, etc.

Further, there are standard ways to produce a new sort of object from an old sort, which can be uniformly applied to all or at least many types: for example, if σ is a sort and τ is a sort, we can talk about collections of σ's or τ's, functions from σ's to τ's, ordered pairs of a σ and a τ, and so forth. These are called *type constructors* when they are considered in general.

In much of this chapter, every variable we introduce will have a type, and a quantifier over that variable will be implicitly restricted to that type.

Sections 2.1.1 and 2.1.2 give an explanation for the typed theory presented in the following sections in an unsorted language with unrestricted quantifiers. This is labelled as a digression from the main development, but it might be useful as an alternative motivation of the system of type theory given below in the main line of the development.

2.1.1 (digression) Unsorted Preamble

We begin with an unsorted theory (in which quantifiers range over all objects), which is a theory of sets, though it is not the usual untyped theory of sets.

We introduce the membership relation $\in$ as a primitive.

In the context of our unsorted set theory, we can represent the notion of being the same type. We certainly expect that all elements of a set will be of the same kind (type). We further expect that for any kind of object, there will be a set of all objects of that kind. So being of the same type will coincide precisely with belonging to some common set.

Definition: We say that x and y are of the same type, written $x \sim_\tau y$, iff $(\exists z : x \in z \wedge y \in z)$.

The terminology suggests that $\sim_\tau$ is an equivalence relation. That this relation is symmetric is a truth of logic. We provide an axiom which makes it easy to demonstrate that it is reflexive and transitive.

Axiom of types: $(\forall x : (\exists y : x \in y \wedge (\forall z : z \in y \leftrightarrow z \sim_\tau x)))$. For each x, we define $\tau(x)$ as the set provided by this axiom which contains x and contains all objects of the same type as x. We call this set the type of x.

$\sim_\tau$ is an equivalence relation: That $x \sim_\tau y \leftrightarrow y \sim_\tau x$ is a tautology of first order logic. From $x \in \tau(x)$ it follows immediately that $x \sim_\tau x$. Suppose $x \sim_\tau y$ and $y \sim_\tau z$. Then $z \sim_\tau y$ (symmetry) so $x \in \tau(y) \wedge z \in \tau(y)$, so $(\exists u : x \in u \wedge y \in u)$, so $x \sim_\tau z$.

We provide that objects with elements are sets, and that sets with the same elements and of the same type are equal.

Axiom of the empty set: We introduce a primitive construction of an object $\emptyset_x$ for each object x, and the axiom $(\forall xy : \emptyset_x \sim_\tau \tau(x) \wedge y \notin x)$. We refer to $\emptyset_x$ as the empty set over x or the empty set of the same type as $\tau(x)$.

Definition of sethood: We define $\texttt{set}(x)$ (x is a set) as

$$(\exists y : y \in x \vee x = \emptyset_y).$$

Axiom of (weak) extensionality:

$$(\forall xy : \mathtt{set}(x) \wedge \mathtt{set}(y) \wedge x \sim_\tau y \wedge (\forall z : z \in x \leftrightarrow z \in y) \rightarrow x = y)$$

It has become stylish in foundations of mathematics to assume that all objects are sets (so there would be no more than one object with no elements). We leave open the possibility that there are many objects with no elements, for more than one reason.

We postulate that every property of objects of a particular kind determines a set.

Axiom scheme of comprehension: For each formula ϕ in which A is not free, we have an axiom

$$(\forall x : (\exists A : \mathtt{set}(A) \wedge A \sim_\tau \tau(x) \wedge (\forall y : y \in A \leftrightarrow x \sim_\tau y \wedge \phi))),$$

For each x, the witness A (unique by extensionality) is denoted by $\{y \sim_\tau x : \phi\}$.

Notational observations: We note that $\emptyset_x$ (the empty set over the type of x) can be expressed as $\{y \sim_\tau x : y \neq y\}$. We note that $\tau(x)$ (the type of x) can be written as

$$\{y \sim_\tau x : y = y\},$$

but the axiom of types does have additional content, since it ensures that x is in this set and so that it is nonempty. We may also write this V_x (the universe containing x). We define $\{y \in \tau(x) : \phi\}$ as $\{y \sim_\tau x : \phi\}$.

We provide a further axiom regulating types. If two nonempty sets are of the same type, their respective elements are of the same type as well.

Axiom of levels: For any x, y, z, w, if $x \sim_\tau y$ and $z \in x$ and $w \in y$, then $z \sim_\tau w$.

Exercise: Show that in the presence of the other axioms, the axiom of levels is equivalent to each of the following assertions:

the axiom of union:

$$(\forall A : \exists U : \mathtt{set}(U) \wedge (\forall x : x \in U \leftrightarrow (\exists a : a \in A \wedge x \in a))).$$

If $A \in \tau^3(x)$ and $\mathtt{set}(A)$, there is a unique such U in $\tau^2(x)$, which we denote by $\bigcup A$.

the axiom of binary union:

$$(\forall AB : \mathtt{set}(A) \wedge \mathtt{set}(B) \wedge A \sim_\tau B \rightarrow (\exists U : \mathtt{set}(U) \wedge (\forall x : x \in U \leftrightarrow x \in A \vee x \in B))).$$

For any particular sets A, B of the same type, we introduce the notation $A \cup B$ for the witness U to this assertion.

Exercise: Verify that the axioms of types and levels both follow if one postulates the existence of $\{x\} = \{y : y = x\}$ for each x (axiom of singletons), the existence of $\mathcal{B}(x) = \{y : x \in y\}$ for each x (the notation commemorates Maurice Boffa's interest in this construction) and the axiom of union.

We consider notation for hierarchies of types.

Definition: For any function symbol F, we define $F^0(x) = x$ and $F^{n+1}(x) = F(F^n(x))$, for each numeral n. Notice that these superscripts are not variables in our language and will not be quantified over. In particular, we now have definitions for $\tau^n(x)$ for any x.

Observations and Definition: Notice that $x \in \tau(x)$ for any x, so $\tau^n(x) \in \tau^{n+1}(x)$. Note further that if $\tau^2(x) = \tau^2(y)$, we have $\tau(x) \sim_\tau \tau(y)$ and with $x \in \tau(x), y \in \tau(y)$ the axiom of levels gives us $x \sim_\tau y$, so $\tau(x) = \tau(y)$. Thus we can define τ^{-1} by $\tau^{-1}(\tau^2(x)) = \tau(x)$, with τ^{-1} defined only at types of types. We can define $\tau^{-n}(x)$ as $(\tau^{-1})^n(x)$, when this is defined.

Note that for any A, $\tau^{-1}(\tau(A))$ is defined iff $\tau(A) = \tau^2(x)$ for some x and in this case is $\tau(x)$, the type of elements of A if A is nonempty, and the type of the same type as A if A is empty. If A is not of a type $\tau^2(x)$ this is undefined.

Definition (common set builder notations): We define $\{x \in A : \phi\}$ for a set A as $\{x \in \tau^{-1}(\tau(A)) : x \in A \wedge \phi\}$: it is indeed our intention that this be defined only if A belongs to some type $\tau^2(x)$. We define

$$\{F(x_1, \ldots, x_n) \in A : \phi\}$$

as

$$\{y \in A : (\exists x_1, \ldots, x_n : y = F(x_1, \ldots, x_n) \wedge \phi)\},$$

where F is any n-ary function symbol (it may be a complex term construction).

The axiom scheme of separation of Zermelo, adorned with technicalities about sethood and type, asserting for each formula ϕ

$$(\forall A : \mathtt{set}(A) \to (\exists B : \mathtt{set}(B) \wedge B \sim_\tau A \wedge (\forall x : x \in B \leftrightarrow x \in A \wedge \phi))),$$

is equivalent to the axiom of comprehension we have presented (in the presence of the other axioms).

Definition (singletons): We define $\iota(x)$ or $\{x\}$ as $\{y \sim_\tau x : y = x\}$.

Definition (unordered pair): If $x \sim_\tau y$, we define $\{x, y\}$ as

$$\{z \sim_\tau x : z = x \vee z = y\}.$$

Note that the existence of the unordered pair of two objects is equivalent to the two objects being of the same type:

$$x \sim_\tau y \leftrightarrow (\exists p : (\forall z : z \in p \leftrightarrow z = x \vee z = y))$$

is a theorem of our system.

We know, because we have shown that $\sim_\tau$ is an equivalence relation, that if two types meet, they are the same. We argue that the types in the hierarchies we have defined are distinct. The argument is related to the paradox of Russell.

Theorem: $\tau(x) \neq \tau^2(x)$.

Proof of Theorem: Suppose otherwise. Then we can define

$$R = \{y \in \tau(x) : y \notin y\},$$

and we will have $R \sim_\tau \tau(x)$ so $R \in \tau^2(x) = \tau(x)$. Then

$$R \in R \leftrightarrow R \in \tau(x) \wedge R \notin R,$$

a contradiction.

This can be generalized.

Theorem: $\tau(x) \neq \tau^{n+2}(x)$.

Proof of Theorem: Suppose otherwise. Note that $\tau(\iota^n(x)) = \tau^{n+1}(x)$. Then we can define

$$R_n = \{\iota^n(y) \in \tau(x) : \iota^n(y) \notin y\},$$

and we will have $R_n \sim_\tau \tau^{n+1}(x)$ so $R_n \in \tau^{n+2}(x) = \tau(x)$ and also $\iota^n(R_n) \in \tau^{n+1}(\tau(x)) = \tau^{n+2}(x) = \tau(x)$. Then

$$\iota^n(R_n) \in R_n \leftrightarrow \iota^n(R_n) \in \tau(x) \wedge \iota^n(R_n) \notin R_n,$$

a contradiction (since $\iota^n(R_n) \in \tau(x)$ is supposed true).

This establishes that for any particular x, the sequence of types $\tau(x), \tau^2(x), \tau^3(x), \ldots$ are distinct and therefore disjoint. This does have one corollary which may feel unusual. The empty set $\emptyset_x \in \tau^2(x)$, and generally $\emptyset_{\tau^n(x)} \in \tau^{n+2}(x)$, so for distinct values of n these empty sets are distinct. This allows us to flesh out our remark that there is more than one reason that we do not adopt strong extensionality: we want to support the possibility of the existence of many atoms which are not sets in each type, and we want to support the existence of distinct empty sets in distinct types.

2.1.2 (digression) Typed language introduced

We explain the type theory we use below using the unsorted language of the previous subsection.

We fix a type $\tau(\mathbf{x})$ ($\mathbf{x}$ being a constant object about which we have very little to say) and define "type n" as $\tau^{n+1}(\mathbf{x})$. We adopt the convention that

each variable x in our language has a natural number type $\mathtt{type}(x)$, and stands for an object in $\tau^{\mathtt{type}(x)+1}(\mathbf{x})$. We read any quantifier $(\forall x : \phi)$ or $(\exists x : \phi)$ as $(\forall x \in \tau^{\mathtt{type}(x)+1}(\mathbf{x}) : \phi)$ or $(\exists x \in \tau^{\mathtt{type}(x)+1}(\mathbf{x}) : \phi)$, respectively. We further adopt the convention that we will only write $x = y$ when $\mathtt{type}(x) = \mathtt{type}(y)$, and we will only write $x \in y$ when $\mathtt{type} + 1 = \mathtt{type}(y)$ (notice that if we wrote an atomic sentence not satisfying the appropriate one of these conditions, it would be false). Notice that in this stereotyped language we cannot even write down the definitions of the sets R and R_n appearing in the proofs that the types are disjoint. We can write the definition of $\sim_\tau$, but as interpreted, any instance of $x \sim_\tau y$ that we can write down consistent with our convention will be true, so this relation does not need to be mentioned.

We can then formulate typed versions of our axioms (which are presented below). It might seem that we are restricting our means of expression by regimenting our language in this way, but this is provably not the case. Any instance of any of the axioms other than comprehension actually satisfies our typed variable conventions. We argue that every instance of comprehension "$\{x \in \tau^n(\mathbf{x}) : \phi\}$ exists" is a consequence of an instance of comprehension which can be expressed in our typed language (so the full unsorted version of the comprehension axiom does not entail the existence of any sets, at least in types $\tau^n(\mathbf{x})$, whose existence is not entailed by the typed axiom scheme).

Any sentence in which no variable of a type $\tau^n(\mathbf{x})$ appears can be replaced with a truth value (easily represented as $u = u$ or its negation): its truth value won't depend on the value of the binding variable x. Each quantified subformula $(\forall y : \psi)$ should be replaced with a conjuncton of versions in which y is assigned a type: all types should be used which are obtained from the type of a parameter or $\tau^n(\mathbf{x})$ by applying $\tau^{\pm(i+1)}$ where i is less than (say) the number of variables in the formula plus one. Similarly, existentially quantified formulas should be replaced with disjunctions of existential formulas restricted to types. When this is done, every variable in the expanded formula will have a type: it is possible that some negative indexed types will be conjured into being. All atomic formulas which do not satisfy the type conventions can be replaced with $\neg u = u$. Logical identities can be used to ensure that no variable not of a type $\tau^{\pm n}(\mathbf{x})$ appears in a quantified formula restricted to a type type $\tau^{\pm n}(\mathbf{x})$, and no variable of this form appears in a quantified formula restricted to a type not of this form: this is done by using logical identities which pull formulas in which the bound variable does not occur out of a quantified formula. Then every formula in which no variable of a type $\tau^{\pm n}(\mathbf{x})$ appears can be replaced with a truth value (either

one, we do not care; either truth value is supported by an instance of typed comprehension) and every formula in which variables of type $\tau^{\pm n}(\mathbf{x})$ appear will actually be well-typed. This transforms our instance of general untyped comprehension into an instance of typed comprehension (or several instances, one of which gives the desired set).

The point of this is the only untyped comprehension axioms that do any work here are the definitions of R and R_n, which manifest themselves only in our assurance that badly typed atomic sentences are false.

It is worth noticing that the unsorted theory may have lots of types other than the types $\tau^{n+1}(\mathbf{x})$ which we use in our typing scheme. The normal expectation that we have types indexed by the natural numbers cannot be conveniently expressed in the unsorted language used here.

2.2 Typed Theory of Sets

We introduce a typed theory of sets in this section, loosely based on the historical type theory of Bertrand Russell. This theory is sufficiently general to allow the construction of all objects considered in classical mathematics. We will demonstrate this by carrying out some constructions of familiar mathematical systems. An advantage of using this type theory is that the constructions we introduce will not be the same as those you might have seen in other contexts, which will encourage careful attention to the constructions and proofs, which furthers other parts of our implicit agenda. Later we will introduce a more familiar kind of set theory.

Suppose we are given some sort of mathematical object (natural numbers, for example). Then it is natural to consider collections of natural numbers as another sort of object. Similarly, when we are given real numbers as a sort of object, our attention may pass to collections of real numbers as another sort of object.

Our approach is an abstraction from this. The basic idea (which we will tweak) is that we introduce a sort of object which we will call *individuals* about which we initially assume nothing whatsoever (we will add an axiom asserting that there are infinitely many individuals when we see how to say this). We also call the sort of individuals *type 0*. We then define *type 1* as the sort of collections of individuals, *type 2* as the sort of collections of type 1 objects, and so forth. (The tweak is that we actually leave open the possibility that each type $n+1$ contains additional objects over and above

the collections of type n objects).

No essential role is played here by natural numbers: we could call type 0 ι and for any type τ let τ^+ be the sort of collections of type τ objects, and then the types 0,1,2... would be denoted $\iota, \iota^+, \iota^{++}, \ldots$ in which we can see that no reference to natural numbers is involved. This paragraph is an answer in advance to an objection raised by philosophers: later we will define the natural number 3 (for example) in type theory: we have not assumed that we already understand what 3 is by using "3" as a formal name for the type ι^{+++}.

Every variable x comes equipped with a type. We may write $x^{\mathbf{3}}$ for a type 3 variable, but we will not always do this: we may write x and expect the type to be deduced from context (type superscripts will be boldface when they do appear so as not to be confused with exponents or other numerical superscripts: we tend to use boldface numerals when we want to emphasize that the use of a numeral in the context under consideration is not a reference to the natural number as a mathematical object). Atomic formulas of our language are of the form $x = y$, in which the variables x and y must be of the same type, and $x \in y$ in which the type of y must be the successor of the type of x.[1]

[1]Just for fun we give a formal description of the grammatical requirements for formulas which does not use numerals (in fact, amusingly, it does not even mention types!). Please note that we will not actually *use* the notation outlined in this paragraph: the point is that the notation we actually use could be taken as an abbreviation for this notation, which makes the point firmly that we are not actually assuming that we know anything about natural numbers yet when we use numerals as type superscripts. We use a more long-winded notation for variables. We make the following stipulations: $\mathbf{x}$ is an individual variable; if y is an individual variable, so is y'; these two rules ensure that we have infinitely many distinct individual (type 0, but we aren't mentioning numerals) variables. Now we define variables in general: an individual variable is a variable; if y is a variable, y^+ is a variable (one type higher, but we are not mentioning numerals). Now we define grammatical atomic formulas. If x is an individual variable and y is a variable, then $x = y$ is an atomic formula iff y is an individual variable. If x is an individual variable, then $x \in y$ is an atomic formula iff y is of the form z^+ where z is an individual variable. For any variables x and y, $x^+ = y^+$ is an atomic formula iff $x = y$ is an atomic formula and $x^+ \in y^+$ is an atomic formula iff $x \in y$ is an atomic formula. We do not write any atomic formula which we cannot show to be grammatical using these rules. The variable consisting of x followed by m primes and n plusses might more conveniently be written $x^{\mathbf{n}}_m$, but in some formal sense it does not have to be: there is no essential reference to numerals here. The rest of the formal definition of formulas: if ϕ is an atomic formula, it is a formula; if ϕ and ψ are formulas and x is a variable, so are (ϕ), $\neg\phi$, $\phi \wedge \psi$, $\phi \vee \psi$, $\phi \rightarrow \psi$, $\phi \leftrightarrow \psi$, $(\forall x.\psi)$, $(\exists x.\psi)$ [interpreting formulas with propositional connectives is

Our theory has axioms. The inhabitants of every type other than 0 are sets (at least, some of them are). We believe that sets are equal iff they have exactly the same elements. This could be expressed as follows:

*__Strong axiom of extensionality:__

$$(\forall x.(\forall y.x = y \leftrightarrow (\forall z.z \in x \leftrightarrow z \in y))),$$

for every assignment of types to x, y, z that makes sense.

*__Proof Strategy:__ If A and B are sets, to prove $A = B$, introduce a new variable a, assume $a \in A$, and deduce $a \in B$, and then introduce a new variable b, assume $b \in B$, and deduce $b \in A$. This strategy simply unfolds the logical structure of the axiom of extensionality.

This axiom says that objects of any positive type are equal iff they have the same elements. This is the natural criterion for equality between sets.

Notice that we did not write

$$(\forall x^{\mathbf{n+1}}.(\forall y^{\mathbf{n+1}}.x^{\mathbf{n+1}} = y^{\mathbf{n+1}} \leftrightarrow (\forall z^{\mathbf{n}}.z^{\mathbf{n}} \in x^{\mathbf{n+1}} \leftrightarrow z^{\mathbf{n}} \in y^{\mathbf{n+1}}))).$$

This would be very cumbersome, and it is not necessary: it is clear from the form of the sentence (it really is a sentence!) that x and y have to have the same type (because $x = y$ appears) and z has to be one level lower in type (because $z \in x$ appears). One does need to be careful when taking this implicit approach to typing to make sure that everything one says *can* be expressed in the more cumbersome notation: we will continue to talk about this where appropriate.

Notice that we starred the strong axiom of extensionality; this is because it is not the axiom we actually adopt. We take the more subtle view that in the real world not all objects are sets [and perhaps not all mathematical constructions are implemented as set constructions], so we might want to allow many non-sets with no elements (it is reasonable to suppose that anything with an element *is* a set). Among the objects with no elements, we designate a particular object $\emptyset$ as the empty *set*.

This does mean that we are making our picture of the hierarchy of types less precise (the tweak that we foreshadowed): type $n + 1$ is inhabited by

made more complicated by order of operations, but the details are best left to a computer parser!].

collections of type n objects and also possibly by other junk of an unspecified nature. A more abstract way of putting it is that our type constructor sending each type τ to a type $\tau+$ is underspecified: all we say is that type $\tau+$ includes the collections of type τ objects.

Primitive notion: There is a designated object $\emptyset^{\mathbf{n+1}}$ for each positive type $n+1$ called the *empty set* of type $n+1$. We do not always write the type index. [2]

Axiom of the empty set: $(\forall x.x \notin \emptyset)$, for all assignments of a type to x and $\emptyset$ which make sense.

Definition: We say that an object x (in a positive type) is a *set* iff

$$x = \emptyset \vee (\exists y.y \in x).$$

We write $\mathtt{set}(x)$ to abbreviate "x is a set" in formulas. We say that objects which are not sets are *atoms* or *urelements*: notice that this only makes sense for objects of positive type. We do not say that individuals (type 0 objects) are atoms, nor do we say that they are not atoms.

Axiom of extensionality:

$$(\forall xy.\mathtt{set}(x) \wedge \mathtt{set}(y) \rightarrow (x = y \leftrightarrow (\forall z.z \in x \leftrightarrow z \in y))),$$

for any assignment of types to variables that makes sense.

Proof Strategy: If A and B are sets, to prove $A = B$, introduce a new variable a, assume $a \in A$, and deduce $a \in B$, and then introduce a new variable b, assume $b \in B$, and deduce $b \in A$. This strategy simply unfolds the logical structure of the axiom of extensionality.

We have already stated a philosophical reason for using a weaker form of the axiom of extensionality, though it may not be clear that this is applicable to the context of type theory (one might at first glance suppose that non-sets are all of type 0); we will see mathematical reasons for adopting the weaker

[2] If we are using the Hilbert symbol in our logic, $(\epsilon x^{\mathbf{n}} : A[x^{\mathbf{n}}])^{\mathbf{n}}$ is of the same type as $x^{\mathbf{n}}$, and $(\epsilon x^{\mathbf{n+1}} : x^{\mathbf{n+1}} \neq x^{\mathbf{n+1}})$ [the default object $\delta^{\mathbf{n+1}}$ of the same type as $x^{\mathbf{n+1}}$] is taken to be $\emptyset^{\mathbf{n+1}}$. The temptation to use $\emptyset^{\mathbf{0}}$ for $\delta^{\mathbf{0}}$ is noted.

form of extensionality in the course of our development (and we will also see mathematical advantages of strong extensionality).

We have said when sets are equal. Now we ask what sets there are. The natural idea is that any property of type n objects should determine a set of type $n+1$, and this is what we will say:

Axiom of comprehension: For any formula $A[x]$ in which the variable y (of type one higher than x) does not appear,

$$(\exists y.\mathtt{set}(y) \wedge (\forall x.x \in y \leftrightarrow A[x]))$$

is an axiom.

This says that for any formula $A[x]$ expressing a property of an object x (of some type n), there is a set y of type $n+1$ such that the elements of y are exactly the objects x such that $A[x]$.

The axiom of extensionality tells us that there is only one such object y which is a set (there may be many such objects y if $A[x]$ is not true for any x, but only one of them ($\emptyset$) will be a set). This suggests a definition:

Set builder notation: For any formula $A[x]$, define $\{x \mid A[x]\}$[3] as the unique *set* of all x such that $A[x]$: an object with exactly these members exists by Comprehension and there is only one such object which is a set by Extensionality. If x is of type n, then $\{x \mid A[x]\}$ is of type $n+1$.[4]

Proof Strategy: To use a posit or deduce a goal of the form $t \in \{x \mid A[x]\}$, replace the posit or goal with the equivalent $A[t]$.

[3]Nowadays (2021) I usually write $\{x : A[x]\}$, and this form may appear in later ecits of the text.

[4]If the Hilbert symbol is used in our logic, we have two comments. First, $\{x : A[x]\}$ can be defined as $(\theta x : (\forall y : y \in x \leftrightarrow A[x]))$; in the case where there might not be a unique object with this extension, the correct one will be chosen magically because the empty set is our default object. Second, one may want to forbid Hilbert symbols appearing in formulas $A[x]$ used in the Axiom of Comprehension, as this amounts to assuming the Axiom of Choice, of which more below. Of course, we *do* as a rule assume the Axiom of Choice, and one might choose to introduce it in this devious way.

A further perhaps amusing observation is that $(\theta x : \mathtt{set}(x) \wedge A[x])$ can be defined as $\{y : (\exists! x.\mathtt{set}(x) \wedge A[x]) \wedge (\forall x : \mathtt{set}(x) \wedge A[x] \rightarrow y \in x)\}$, without any assumption that the Hilbert symbol is in use.

In our numeral free notation we indicate the grammar requirements for set abstracts: if x is a variable and ϕ is a formula, $\{x \mid \phi\}$ can replace any occurrence of x^+ in a formula and it will still be a formula.

There are two other axioms in our system, the Axiom of Infinity and the Axiom of Choice, but some formal development should be carried out before we introduce them.

2.3 Russell's Paradox?

At this point an objection might interpose itself. Consider the following argument.

For any set x, obviously either x is an element of itself or x is not an element of itself. Form the set R whose elements are exactly those sets which are not elements of themselves: $R = \{x \mid x \notin x\}$. Now we ask, is R an element of itself? For any x, $x \in R \leftrightarrow x \notin x$, so in particular $R \in R \leftrightarrow R \notin R$. This is a contradiction!

This argument, known as *Russell's paradox*, was a considerable embarrassment to early efforts to formalize mathematics on the very abstract level to which we are ascending here.

Fortunately, it is completely irrelevant to our work here. This argument does not work in our system, on a purely formal level, because $x \in x$ is not a legal formula in the language of our type theory, so it does not define a property of sets allowing the introduction of a set by Comprehension! On a less formal level, attending to the meaning of notations rather than their formal structure, we have not introduced the kind of sweeping notion of set presupposed in the argument for Russell's paradox: for any particular sort of object τ (such as type n) we have introduced the new sort of object "set of τ's" or "τ^+" (which we call type $n+1$ in the particular case where τ is type n). The supposition in Russell's paradox is that we have a type of sets which contains all sets of objects of that same type. Ordinary mathematical constructions do not lead us to a situation where we need such a type. If we had a universal sort o containing *all objects* it might seem that o^+ would contain all sets of anything whatsoever (including sets of type o^+ sets, which would presumably also be of the universal type o). The argument for Russell's paradox shows that there cannot be such a type if the Axiom of Comprehension is to apply: either there cannot be a universal type o or the type o^+ cannot contain all definable subcollections of o. We will introduce untyped set theories with restrictions on comprehension below.

It is important to notice on a philosophical level that care in the introduction of the idea of a set has completely avoided the paradox: there is no embarrassment for our typed notion of set, and our typed notion of set is true to what we actually do in mathematics. Russell's paradox was a serious problem for an initial insufficiently careful development of the foundation of mathematics; it is not actually a problem for the foundations of mathematics as such, because the typed notion of set is all that actually occurs in mathe-

matics in practice (in spite of the fact that the system of set theory which is customarily used is formally untyped: we shall meet this system in chapter 3 and see that its restrictions on comprehension can be naturally motivated in terms of types).

Notice that if x and y are terms of different types, $x = y$ is not a formula at all. This does not mean that we say that x and y are distinct: it means that we do not entertain the question as to whether objects of different types are identical or distinct (for now; we will have occasion to think about this later). Similarly, if the type of y is not the successor of the type of x (for example, if x and y are of the same type) we do not say $x \in y$ (it is ungrammatical, not false). We do not ask whether $x \in x$; we do not say that it is false (or true) (for now).

If the reader has looked at sections 2.1.1 and 2.1.2, she will have seen an application of the Russell argument to show a fact we cannot express in our typed language.

2.4 Simple Ideas of Set Theory

In this section we develop some familiar ideas of set theory.

We first develop the familiar list notation for finite sets. Here are the standard notations for one and two element sets.

List notation for sets: $\{x\}$ is defined as $\{y \mid y = x\}$. $\{x, y\}$ is defined as $\{z \mid z = x \vee z = y\}$.

It is convenient to define Boolean union and intersection of sets before giving the general definition of list notation.

Boolean union and intersection: If x and y are sets, define $x \cup y$ as

$$\{z \mid z \in x \vee z \in y\}$$

and $x \cap y$ as

$$\{z \mid z \in x \wedge z \in y\}.$$

Notice that though we may informally think of $x \cup y$ as "x and y", it is actually the case that $x \cup y$ is associated with the logical connective $\vee$ and it is $x \cap y$ that is associated with $\wedge$ in a logical sense.

We also define a^c (the complement of a) as $\{x \mid x \notin a\}$ and $a - b$ (the set difference of a and b) as $a \cap b^c$.

recursive definition of list notation: $\{x_1, x_2, \ldots, x_n\}$ is defined as

$$\{x_1\} \cup \{x_2, \ldots, x_n\}.$$

Notice that the definition of list notation for n items presupposes the definition of list notation for $n - 1$ items: since we have a definition of list notation for 1 and 2 items we have a basis for this recursion.

Note that all elements of a set defined by listing must be of the same type, just as with any set.

There is one more very special case of finite sets which needs special attention.

null set: We have introduced $\emptyset^{\mathbf{n+1}}$ as a primitive notion because we adopted the weak axiom of extensionality.

If we assumed strong extensionality, we could define $\emptyset^{\mathbf{n+1}}$ as

$$\{x^{\mathbf{n}} \mid x^{\mathbf{n}} \neq x^{\mathbf{n}}\}$$

(in any event this set abstract is equal to $\emptyset^{\mathbf{n+1}}$!). Notice that $\emptyset^{\mathbf{n+1}}$ has no elements, and it is by Extensionality (either form) the only set (of type $n+1$) with no elements. In this definition we have used type superscripts, though hereinafter we will write just $\emptyset$: this is to emphasize that $\emptyset$ is defined in each positive type and we do not say that the empty sets in different types are the same (or that they are different). Notice that although $x \in x$ is not grammatical, $\emptyset \in \emptyset$ *is* grammatical (and false!). It is not an instance of the ungrammatical form $x \in x$ because the apparent identity of the two occurrences of $\emptyset$ is a kind of pun. The pun can be dispelled by writing $\emptyset^{\mathbf{n+1}} \in \emptyset^{\mathbf{n+2}}$ explicitly.

universe: We define V as $\{x \mid x = x\}$. This is the universal set. The universal set in type $n+1$ is the set of all type n objects. $V \in V$ is grammatical and true – but the two occurrences of V have different reference (this can be written $V^{\mathbf{n+1}} \in V^{\mathbf{n+2}}$ for clarification).

Of course we assume that the universal set is not finite, but we do not know how to say this yet.

The combination of the empty set and list notation allows us to write things like $\{\emptyset, \{\emptyset\}\}$, but not things like $\{x, \{x\}\}$: the former expression is another pun, with empty sets of different types appearing, and the latter expression is ungrammatical, because it is impossible to make a consistent type assignment to x. An expression like this can make sense in an untyped set theory (and in fact in the usual set theory the first expression here is the most popular way to define the natural number 2, as we will explain later).

Set builder notation can be generalized.

Generalized set builder notation: If we have a complex term $t[x_1, \ldots, x_n]$ containing only the indicated variables, we define $\{t[x_1, \ldots, x_n] \mid A\}$ as $\{y \mid (\exists x_1 \ldots x_n . y = t[x_1, \ldots, x_n] \wedge A)\}$ (where y is a new variable). We do know that this kind of very abstract definition is not really intelligible in practice except by backward reference from examples, and we will provide these!

Examples: $\{\{x\} \mid x = x\}$ means, by the above convention,

$$\{z \mid (\exists z.z = \{x\} \wedge x = x)\}.$$

It is straightforward to establish that this is the set of all sets with exactly one element, and we will see below that we will call this the natural number 1. The notation $\{\{x, y\} \mid x \neq y\}$ expands out to $\{z \mid (\exists xy.z = \{x, y\} \wedge x \neq y)\}$: this can be seen to be the set of all sets with exactly two elements, and we will identify this set with the natural number 2 below.

We define some familiar relations on sets.

subset, superset: We define $A \subseteq B$ as

$$\mathtt{set}(A) \wedge \mathtt{set}(B) \wedge (\forall x.x \in A \rightarrow x \in B).$$

We define $A \supseteq B$ as $B \subseteq A$.

Theorem: For any set A, $A \subseteq A$.

Theorem: For any sets A, B, $A \subseteq B \wedge B \subseteq A \rightarrow A = B$.

Theorem: For any sets A, B, C, if $A \subseteq B$ and $B \subseteq C$ then $A \subseteq C$.

Observation: The theorems we have just noted will shortly be seen to establish that the subset relation is a "partial order".

Proof Strategy: To show that $A \subseteq B$, where A and B are known to be sets, introduce an arbitrary object x and assume $x \in A$: show that it follows that $x \in B$.

If one has a hypothesis or previously proved statement $A \subseteq B$ and a statement $t \in A$, deduce $t \in B$.

Notice that the proof strategy given above for proving $A = B$ is equivalent to first proving $A \subseteq B$, then proving $B \subseteq A$.

The notions of element and subset can be confused, particularly because mathematicians and math students have a bad habit of saying things like "A is in B" or "A is contained in B" both for $A \in B$ and for $A \subseteq B$. It is useful to observe that elements are not "parts" of sets. The relation of part

to whole is transitive: if A is a part of B and B is a part of C, then A is a part of C. The membership "relation" is not transitive in a quite severe sense: if $A \in B$ and $B \in C$, then $A \in C$ is not even meaningful in our type theory! [In the untyped set theories discussed in chapter 3, membership is in a quite normal sense not transitive.] But the subset relation is transitive: if $A \subseteq B$ and $B \subseteq C$, then any element of A is also an element of B, and so is in turn an element of C, so $A \subseteq C$. If a set can be said to have parts, they will be its subsets, and its one-element sets $\{a\}$ for $a \in A$ can be said to be its atomic parts.

We give a general format for introducing operations, and then introduce an important operation.

Definable Operations: For any formula $\phi[x, y]$ with the property that

$$(\forall xyz.\phi[x, y] \wedge \phi[x, z] \rightarrow y = z)$$

we define $F_\phi(x)$ or F_ϕ‘x as the unique y (if there is one) such that $\phi[x, y]$. Note that we will not always explicitly give a formula ϕ defining an operation, but it should always be clear that such a formula could be given. Note also that there might be a type differential between x and $F_\phi(x)$ depending on the structure of the formula $\phi[x, y]$.

For any such definable operation $F(x)$, we define F“x for any set x as $\{F(u) \mid u \in x\}$: F“x is called the (elementwise) *image* of x under the operation F.

We also support iteration of such operations: $F^{\mathbf{0}}(x)$ is defined as x and $F^{\mathbf{n+1}}(x)$ is defined as $F(F^{\mathbf{n}}(x))$. The numerals here are in boldface to indicate that no reference to natural numbers as mathematical objects is intended.

Power Set: For any set A, we define $\mathcal{P}(A)$ as $\{B \mid B \subseteq A\}$. The power set of A is the set of all subsets of A. Notice that $\mathcal{P}(V^{\mathbf{n}})$ is the collection of all sets of type $n+1$, and is not necessarily the universe $V^{\mathbf{n+1}}$, which might also contain some atoms.

Singleton: For any object x, we define $\iota(x) = \{x\}$. The primary use of this alternative notation for the singleton operation is to allow notations like $\iota^{\mathbf{3}}(x)$ for $\{\{\{x\}\}\}$.

Observation: $\mathcal{P}$ is F_ϕ where $\phi[x, y]$ is the formula $(\forall z.z \in y \leftrightarrow z \subseteq x)$ (or just $y = \{z \mid z \subseteq x\}$). The operator ι is F_ϕ where ϕ is $(\forall z.z \in y \leftrightarrow z = x)$.

It is **very important** to notice that $\mathcal{P}(x)$ is one type higher than x, and similarly that $\iota(x) = \{x\}$ is one type higher than x.

It may be well known to you that union, intersection and complement satisfy the following properties (which demonstrate that the sets in each type with these operations form what is called a *Boolean algebra*). You should also notice that these are closely parallel with the properties of disjunction, conjuction, and negation, the logical operations which appear in the definitions of the set operations.

commutative	$A \cup B = B \cup A$		$A \cap B = B \cap A$
associative	$(A \cup B) \cup C = A \cup (B \cup C)$		$(A \cap B) \cup C = A \cap (B \cap C)$
identity	$A \cup \emptyset = A$		$A \cap V = A$
zero	$A \cup V = V$		$A \cap \emptyset = \emptyset$
idempotent	$A \cup A = A$		$A \cap A = A$
distributive	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$		$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
cancellation		$(A^c)^c = A$	
deMorgan	$(A \cup B)^c = A^c \cap B^c$		$(A \cap B)^c = A^c \cup B^c$

The properties motivate a style in which $A \cup B$ (or $A \vee B$) is written $a + b$, $A \cap B$ (or $A \wedge B$) is written ab, V (or `true`) is written 1 and $\emptyset$ (or `false`) is written 0. The complement (or negation) operation, which doesn't really correspond to anything in arithmetic, is often written with an overline: $\overline{a}$ represents A^c (or $\neg A$).

On the next page, we give a proof of a sample axiom of Boolean algebra. This is tedious in obvious ways!

Prove: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$. The objects to be shown equal are obviously sets: we use the set equality strategy.

Part I:

Assume (1): $x \in A \cup (B \cap C)$

Goal: $x \in (A \cup B) \cap (A \cup C)$

(2): $x \in A \vee x \in B \cap C$ definition of union, 1

We prove the result by cases on 2

Case I: assume (2a): $x \in A$

Goal: $x \in (A \cup B) \cap (A \cup C)$

(3): $x \in A \vee x \in B$ addition (2a)

(4): $x \in A \vee x \in C$ addition ((2a)

(5): $x \in A \cup B$ def union 3

(6): $x \in A \cup C$ def union 4

(6.5) $x \in (A \cup B) \wedge x \in (A \cup C)$ conj 5,6

(7): $x \in (A \cup B) \cap (A \cup C)$ def intersection 6.5

Case II: assume (2a): $x \in B \cap C$

Goal: $x \in (A \cup B) \cap (A \cup C)$

(8): $x \in B \wedge x \in C$ def intersection 2a

(9): $x \in B$ simp 8

(10): $x \in C$ simp 8

(11): $x \in A \vee x \in B$ addition 9

(12): $x \in A \vee x \in C$ addition 10

(13): $x \in A \cup B$ def union 11

(14): $x \in A \cup C$ def union 12

(15): $x \in (A \cup B) \wedge x \in (A \cup C)$ conj 13,14

(16): $x \in (A \cup B) \cap (A \cup C)$ def intersection 15

(17): $x \in (A \cup B) \cap (A \cup C)$ proof by cases, 2, 2a–7, 2b–16

Part II:

Assume (18): $x \in (A \cup B) \cap (A \cup C)$

Goal: $x \in A \cup (B \cap C)$

Goal: $x \in A \vee x \in B \cap C$ (rewriting goal using definition of union)

Assume (19): $\neg x \in A$

Goal: $x \in B \cap C$

(20): $x \in A \cup B \wedge x \in A \cup C$ def intersection 17

(21): $x \in A \cup B$ simp 20

(22): $x \in A \cup C$ simp 20

23): $x \in A \vee x \in B$ def union 21

(24): $x \in A \vee x \in C$ def union 22

(25): $x \in B$ d.s. 19, 23

(26): $x \in C$ d.s. 19, 24

(27): $x \in B \wedge x \in C$ conj 25, 26

(28): $x \in B \cap C$ def int 27

(29): $x \in A \vee x \in B \cap C$ alt elim 19-28

(30): $x \in A \cup (B \cap C)$ def union 29

the main result is proved: by set equality strategy, 1-17, 18-30

2.4.1 Review of set theory proof strategies

In the first chapter, you should have noticed our orientation toward letting the form of a statement drive the way we prove it and the way we use it. This approach of identifying proof strategies can be extended to set theory. We present some rules along these lines, some of which will already have been mentioned.

Here is a collection of official basic rules for reasoning about sets in our logic. We assume the type constraints on language stated above, but we don't attach types explicitly to variables here. The basic notions are membership, equality, the sethood predicate, and set builder notation.

You might want to look back to section 1.11 for basic rules of equality (reflexivity and substitution) as well as familiar derived rules.

Rule of nonemptiness: things with elements are sets.

$$\frac{t \in u}{\mathtt{set}(u)}$$

Rule of abstraction: set abstracts (the objects denoted by set builder notation) are sets. No premises are needed: this could also simply be stated as an axiom.

$$\frac{}{\mathtt{set}(\{x : P[x]\})}$$

Rule of extensionality for set abstracts.

$$\frac{(\forall x : P[x] \leftrightarrow Q[x])}{\{x : P[x]\} = \{x : Q[x]\}}$$

General rule of extensionality: the preceding rule can be deduced from this one and the rule of abstraction. The extensionality rules are further expanded into a more complicated proof strategy using the rules for the universal quantifier and the biconditional below.

$$\frac{\mathtt{set}(t) \quad \mathtt{set}(u) \quad (\forall x : x \in t \leftrightarrow x \in u)}{t = u}$$

Leibniz rules for equality: (these could even be taken as a definition of equality). Notice that this is not extensionality: this says that things are equal iff they belong to the same sets. The official rules of equality, reflexivity and substitution, can be deduced from these rules with the aid of other rules of logic and the rules of comprehension below (try it).

$$\frac{(\forall x : t \in x \leftrightarrow u \in x)}{t = u}$$

$$\frac{t = u}{(\forall x : t \in x \leftrightarrow u \in x)}$$

First rule of comprehension

$$\frac{P[t]}{t \in \{x : P[x]\}}$$

Second rule of comprehension

$$\frac{t \in \{x : P[x]\}}{P[t]}$$

Here is a discussion of proof strategies.

to prove that sets are equal: Given that A and B are sets, to prove $A = B$, prove $(\forall x : x \in A \leftrightarrow x \in B)$. This expands out using the proof strategies for the universal quantifier and the biconditional.

Like the proof of a biconditional, the proof is divided into two parts.

Let c be arbitary. Assume $c \in A$ for the sake of argument: show that $c \in B$.

Let d be arbitrary and assume $d \in B$ for the sake of argument: show that $d \in A$.

If you have shown these two things, you have shown $A = B$.

to prove subset relations: Let A, B be sets. To prove $A \subseteq B$ is to prove $(\forall x : x \in A \to x \in B)$.

This expands out to the following strategy:

Let c be arbitary. Assume that $c \in A$. Show that $c \in B$.

If you can carry this out, you have shown $A \subseteq B$.

to prove equality using the subset relation: Notice that if you have shown $A \subseteq B$ and $B \subseteq A$, you have shown $A = B$.

to use a subset statement: If you have $a \in A$ and $A \subseteq B$, deduce $a \in B$.

to prove that an object belongs to a specific set: To show that $a \in \{x : P[x]\}$, prove $P[a]$. This is of course very abstract. Every defined set theoretical operation amounts to a special case of this.

to use an assertion that an object belongs to a specific set: If you have $a \in \{x : P[x]\}$, deduce $P[a]$. Again, this is very abstract, but every defined set theoretical operation gives us a version of this move.

To show that $a \in \{x : P[x]\}$, prove $P[a]$. This is of course very abstract. Every defined set theoretical operation amounts to a special case of this.

a general strategy: when all else fails, expand definitions. To prove or use statements involving defined operations, expand the definitions (and use the comprehension axiom).

For example, to prove $a \in \{b\}$, notice that this is $\{x : x = b\}$, so prove $a = b$.

Similarly, if you have proved or assumed $a \in \{b\}$, you may further deduce $a = b$.

Similarly, $x \in A \cup B$ is equivalent to $x \in A \vee x \in B$, which should be exploited if we want to use or prove a statement about unions.

$x \in A \cap B$ is equivalent to $x \in A \wedge x \in B$, which should be exploited if we want to use or prove a statement about intersections.

$x \in \mathcal{P}(A)$ is equivalent to $x \subseteq A$ (which will then involve opportunities to apply rules for handling the subset relation).

$x \in \bigcup \mathcal{A}$ is equivalent to $(\exists A : x \in A \wedge A \in \mathcal{A})$, which suggests a strategy "introduce a witness, a set $A \in \mathcal{A}$ such that $x \in A$".

complex definitions: Rinse and repeat. For example, if we have to use the assertion $a \in b \cup \{c\}$, we can expand this to $a \in b \vee a = c$. With experience, this should be pretty automatic.

2.4.2 Exercises

1. Prove $A \cap B = B \cap A$ (using the definition of intersection and the proof strategy indicated for equality of sets: it comes down to that proof strategy and very easy propositional logic).

2. Prove $(A \cup B)^c = A^c \cap B^c$ (recalling that A^c, the complement of A is defined as $\{x \mid x \notin A\}$).

3. Prove $(A - B) - C = A - (B \cup C)$.

4. Verify the validity of the rule

$$\frac{t \in A \qquad\qquad A \subseteq B}{t \in B}$$

   using basic rules of logic and the rules for reasoning with sets given in the previous section. You may then use this rule (call it "subset rule").

5. Prove the theorem $A \subseteq B \wedge B \subseteq C \rightarrow A \subseteq C$. Look at the proof strategies for the subset relation in the text. Roughly the same proof justifies the "rule of transitivity of subset",

$$\frac{A \subseteq B \qquad\qquad B \subseteq C}{A \subseteq C}$$

6. Prove $A \subseteq B \rightarrow \mathcal{P}(A) \subseteq \mathcal{P}(B)$. I did this in class in a rough form: so obviously I would like something closer to a line by line proof, and in the text above I have given you finer-grained rules and strategies that you may employ.

7. Give possible assignments of types to variables in each of the following statements or set expressions, or say that they cannot really be statements in our typed language (with at least a brief explanation). There will not be a unique answer, because you can take any assignment of types that works and add one to every type you use (this is an important observation!)

   We give some schematics as a reminder:

$$x^{\mathbf{n}} = y^{\mathbf{n}}; x^{\mathbf{n}} \in y^{\mathbf{n+1}}; \{x^n : P[x^n]\}^{\mathbf{n+1}}.$$

Of course, any defined set operation has some effect on types: for example, $(x^{\mathbf{n}} \bigcup y^{\mathbf{n}})^{\mathbf{n}}$, union doesn't move types, but $\mathcal{P}(x^{\mathbf{n}})^{\mathbf{n+1}}$, the output of the power set operation is one type higher than the input. Problem 7 is intended to give you a little tutorial in thinking about types. I hope I have given enough hints!

(a)
$$(\forall x \in A : (\exists y \in x : y \in B))$$

(b)
$$(\forall x : x \in A \wedge y \in x \rightarrow x \in y)$$

[in ordinary set theory, this makes sense and the set A is said to be "transitive" if this is true. Here it does not. Explain why.]

(c) We attempt to define x^+ as $x \cup \{x\}$. Does this make sense in type theory? If it does, give a possible assignment of types to x and x^+. If it doesn't, explain the problem. Hint: you need to think about the type of $\{x\}$ too.

(d) We define the set $\mathcal{F}$ of finite sets as $\{x : (\forall I : \emptyset \in I \wedge (\forall xy : x \in I \rightarrow x \in \{y\} \in I) \rightarrow x \in I\}$. We assure you that this does indeed make sense. Find a sensible assignment of types to $\mathcal{F}$ and all the variables in its definition.

(e) Compare the similar assertions $\{x : x = x\} \in \{y : y = y\}$ and $\{x : x = x\} \in \{x : x = x\}$. One of these makes sense in type theory and one of them doesn't. Explain.

2.5 Digression: simple ideas of set theory in the language of the unsorted preamble

In this section we develop some familiar ideas of set theory, in the unsorted language of section 2.1.1. This section is here just to give a flavor of what a development without typed language of the same theory might look like.

We first develop the familiar list notation for finite sets. Here are the standard notations for one and two element sets.

List notation for sets: $\{x\}$ is defined as $\{y \sim_\tau x : y = x\}$. $\{x, y\}$ is defined, when $x \sim_\tau y$, as $\{z \sim_\tau x : z = x \vee z = y\}$.

It is convenient to define Boolean union and intersection of sets before giving the general definition of list notation.

Boolean union and intersection: If x and y are sets of the same type, define $x \cup y$ as

$$\{z \in \tau^{-1}(\tau(x)) : z \in x \vee z \in y\}$$

and $x \cap y$ as

$$\{z \in \tau^{-1}(\tau(x)) : z \in x \wedge z \in y\}.$$

Notice that though we may informally think of $x \cup y$ as "x and y", it is actually the case that $x \cup y$ is associated with the logical connective $\vee$ and it is $x \cap y$ that is associated with $\wedge$ in a logical sense.

The axiom of levels plays an important role here, ensuring that the fact that $x \sim_\tau y$ implies that elements of x are of the same type as elements of y.

We also define a^c (the complement of a) as $\{x \in \tau^{-1}(\tau(a)) : x \notin a\}$ and $a - b$ (the set difference of a and b) as $a \cap b^c$ (of course under the assumption $a \sim_\tau b$).

recursive definition of list notation: $\{x_1, x_2, \ldots, x_n\}$ is defined as

$$\{x_1\} \cup \{x_2, \ldots, x_n\}.$$

Notice that the definition of list notation for n items presupposes the definition of list notation for $n - 1$ items: since we have a definition of list notation for 1 and 2 items we have a basis for this recursion.

Note that all elements of a set defined by listing must be of the same type, just as with any set.

There is one more very special case of finite sets which needs special attention.

null set: We define $\emptyset_x$ as $\{y \in \tau(x) : y \neq y\}$. Notice that $\emptyset_x \in \tau^2(x)$. It would be natural to extend our typed language convention to allow $\emptyset$ to be used as a constant without the subscript in situations where its type can be deduced from that of neighboring variables. A statement such as $\emptyset \in \emptyset$ could be read as $\emptyset_x \in \emptyset_{\tau(x)}$ (though strictly speaking probably one of the subscripts should be supplied). In any event, all such statements are false.

universe: We define V_x as $\{y \in \tau(x) : y = y\}(= \tau(x))$. Notice that $V_x \in \tau^2(x)$. It would be natural to extend our typed language convention to allow V to be used as a constant without the subscript in situations where its type can be deduced from that of neighboring variables. A statement such as $V \in V$ could be read as $V_x \in V_{\tau(x)}$ (though strictly speaking probably one of the subscripts should be supplied). In any event, all such statements are true.

Of course we assume that the universal set is not finite, but we do not know how to say this yet.

The combination of the empty set and list notation allows us to write things like $\{\emptyset, \{\emptyset\}\}$, but not things like $\{x, \{x\}\}$: the former expression is another pun, with empty sets of different types appearing, and the latter expression is undefined, because x and $\{x\}$ have different types. An expression like this can make sense in a more usual untyped set theory (and in fact in the usual set theory the first expression here is the most popular way to define the natural number 2, as we will explain later).

Set builder notation can be generalized.

Generalized set builder notation: If we have a complex term $t[x_1, \ldots, x_n]$ containing only the indicated variables, we define

$$\{t[x_1, \ldots, x_n] \in \tau(u) : A\}$$

as $\{y \in \tau(u) : (\exists x_1 \ldots x_n . y = t[x_1, \ldots, x_n] \wedge A)\}$ (where y is a new variable). We do know that this kind of very abstract definition is not really intelligible in practice except by backward reference from examples, and we will provide these!

Examples: $\{\{x\} \in \tau(u) : x = x\}$ means, by the above convention,

$$\{z \in \tau(u) \mid (\exists z.z = \{x\} \wedge x = x)\}.$$

It is straightforward to establish that this is the set of all sets with exactly one element, and we will see below that we will call this the natural number 1. The notation $\{\{x, y\} \in \tau(u) : x \neq y\}$ expands out to $\{z \in \tau(u) \mid (\exists xy.z = \{x, y\} \wedge x \neq y)\}$: this can be seen to be the set of all sets with exactly two elements (belonging to $\tau^2(u)$), and we will identify this set with the natural number 2 below.

We define some familiar relations on sets.

subset, superset: We define $A \subseteq B$ as

$$\mathtt{set}(A) \wedge \mathtt{set}(B) \wedge A \sim_\tau B \wedge (\forall x.x \in A \rightarrow x \in B).$$

We define $A \supseteq B$ as $B \subseteq A$.

Theorem: For any set A, $A \subseteq A$.

Theorem: For any sets A, B, $A \subseteq B \wedge B \subseteq A \rightarrow A = B$.

Theorem: For any sets A, B, C, if $A \subseteq B$ and $B \subseteq C$ then $A \subseteq C$.

Observation: The theorems we have just noted will shortly be seen to establish that the subset relation is a "partial order".

Proof Strategy: To show that $A \subseteq B$, where A and B are known to be sets of the same type (and so with elements of the same type), introduce an arbitrary object x and assume $x \in A$: show that it follows that $x \in B$.

If one has a hypothesis or previously proved statement $A \subseteq B$ and a statement $t \in A$, deduce $t \in B$.

Notice that the proof strategy given above for proving $A = B$ is equivalent to first proving $A \subseteq B$, then proving $B \subseteq A$.

The notions of element and subset can be confused, particularly because mathematicians and math students have a bad habit of saying things like "A is in B" or "A is contained in B" both for $A \in B$ and for $A \subseteq B$. It is useful to observe that elements are not "parts" of sets. The relation of part

to whole is transitive: if A is a part of B and B is a part of C, then A is a part of C. The membership "relation" is not transitive in a quite severe sense: if $A \in B$ and $B \in C$, then $A \notin C$, because $C \in \tau^2(B) = \tau^3(A)$, and any set to which A belongs is an element of $\tau^2(A)$, which is disjoint from $\tau^3(A)$. But the subset relation is transitive: if $A \subseteq B$ and $B \subseteq C$, then any element of A is also an element of B, and so is in turn an element of C, so $A \subseteq C$. If a set can be said to have parts, they will be its subsets, and its one-element sets $\{a\}$ for $a \in A$ can be said to be its atomic parts.

We give a general format for introducing operations, and then introduce an important operation.

Definable Operations: For any formula $\phi[x, y]$ with the property that

$$(\forall xyz.\phi[x, y] \wedge \phi[x, z] \rightarrow y = z)$$

we define $F_\phi(x)$ or F_ϕ‘x as the unique y (if there is one) such that $\phi[x, y]$. Note that we will not always explicitly give a formula ϕ defining an operation, but it should always be clear that such a formula could be given. Note also that there might be a type differential between x and $F_\phi(x)$ depending on the structure of the formula $\phi[x, y]$: we require that $\tau(F_\phi(x))$ be expressible for any x for which it is defined in some form $\tau^m(\tau^{-n}(\tau(x)))$ for fixed m, n.

For any such definable operation $F(x)$, we define F“x for any set x as $\{F(u) \in \tau^m(\tau^{-n}(\tau(x))) : u \in x\}$, where $\tau^m(\tau^{-n}(\tau(x)))$ is the type of values of F with arguments taken from x: F“x is called the (element-wise) *image* of x under the operation F.

We also support iteration of such operations: $F^{\mathbf{0}}(x)$ is defined as x and $F^{\mathbf{n+1}}(x)$ is defined as $F(F^{\mathbf{n}}(x))$. The numerals here are in boldface to indicate that no reference to natural numbers as mathematical objects is intended.

Power Set: For any set A, we define $\mathcal{P}(A)$ as $\{B \in \tau(A) : B \subseteq A\}$. The power set of A is the set of all subsets of A. Notice that $\mathcal{P}(V_x)$ is the collection of all sets in $\tau^2(x)$, and is not necessarily the universe $\tau^2(x)$, which might also contain some atoms.

Singleton: For any object x, we define $\iota(x) = \{x\}$. The primary use of this alternative notation for the singleton operation is to allow notations like $\iota^{\mathbf{3}}(x)$ for $\{\{\{x\}\}\}$.

Observation: $\mathcal{P}$ is F_ϕ where $\phi[x, y]$ is the formula $(\forall z.z \in y \leftrightarrow z \subseteq x)$ (or just $y = \{z \in \tau(x) : z \subseteq x\}$). The operator ι is F_ϕ where ϕ is $(\forall z.z \in y \leftrightarrow z = x)$.

It is **very important** to notice that $\mathcal{P}(x)$ is one type higher than x (belongs to $\tau^2(x)$ instead of $\tau(x)$), and similarly that $\iota(x) = \{x\}$ is one type higher than x.

2.6 Finite Number; the Axiom of Infinity; Ordered Pairs

In the usual untyped set theory, the natural numbers are usually defined using a clever scheme due to John von Neumann.

*__Definition:__ 0 is defined as $\emptyset$. 1 is defined as $\{0\}$. 2 is defined as $\{0, 1\}$. 3 is defined as $\{0, 1, 2\}$. In general, $n + 1$ is defined as $n \cup \{n\}$.

The star on this "definition" indicates that we do not use it here. The problem is that this definition makes no sense in our typed language. Notice that there is no consistent way to assign a type to n in "$n \cup \{n\}$". In chapter 3 on untyped set theory, we will be able to use this definition and we will see that it generalizes to an incredibly slick definition of ordinal number.

The motivation of our definition of natural number in type theory is the following

Circular Definition: The natural number n is the set of all sets with n elements.

Of course this will not be acceptable as a formal definition: we spend the rest of the section showing how we can implement it using a series of formally valid definitions.

It is amusing to observe that the von Neumann definition above can also be motivated using another

*__Circular Definition:__ The natural number n is the set of all natural numbers less than n.

This is starred to indicate that we are not at this point using it at all!

Definition: We define 0 as $\{\emptyset\}$.

Note that we have thus defined 0 as the set of all sets with zero (no) elements.

Definition: For any set A, define $\sigma(A)$ as $\{x \cup \{y\} \mid x \in A \wedge y \notin x\}$. $\sigma(A)$, which we call the *successor* of A, is the collection of all sets obtained by adjoining a single new element to an element of A.

Alternative Definition: For any sets A, B, define $A + B$ as

$$\{a \cup b \mid a \in A \wedge b \in B \wedge a \cap b = \emptyset\}.$$

Define 1 as

$$\{x \mid (\exists y : (\forall z : z \in x \leftrightarrow z = y))\},$$

or equivalently $\{\{x\} : x = x\}$. It is straightforward to see that $A + 1$ defined using these definitions is the same as $\sigma(A)$. This notion of addition for general sets extends the notion we will eventually define for natural numbers.

Definition: We define 1 as above or as $\sigma(0)$. (Observe that 1 is the set of all one-element sets (singletons).) We define 2 as $\sigma(1)$ or 1+1, 3 as $\sigma(2)$ or 2+1, and so forth (and observe that 2 is the set of all sets with exactly two elements, 3 is the set of all sets with exactly three elements, and so forth).

Unfortunately, "and so forth" is a warning that a careful formal examination is needed at this point!

Definition: We call a set I an *inductive set* if $0 \in I$ and

$$(\forall A.A \in I \rightarrow \sigma(A) \in I).$$

We define $\mathcal{I}$ as the set of all inductive sets.

At this point it is useful to define the unions and intersections of not necessarily finite collections of sets.

Definition: For any set A, we define $\bigcup A$ as

$$\{x \mid (\exists a \in A.x \in a)\}$$

and $\bigcap A$ as

$$\{x \mid (\forall a \in A.x \in a)\}.$$

(Notice that $x \cup y = \bigcup\{x, y\}$ and $x \cap y = \bigcap\{x, y\}$.)[5]

Observation: Notice that $\bigcup A$ and $\bigcap A$ are of type $\mathbf{n+1}$ if A is of type $\mathbf{n+2}$ (we are using boldface here to clearly indicate where I am talking about types rather than natural numbers).

Definition: We define $\mathbb{N}$, the set of all natural numbers, as $\bigcap \mathcal{I}$, the intersection of all inductive sets.[6]

We saw above that 0 has been successfully defined as the set of all zero element sets, 1 as the set of all one-element sets, 2 as the set of all two-element sets and so forth (whenever and so forth, etc, ... or similar devices appear in mathematical talk, it is a signal that there is something the author hopes you will see so that he or she does not have to explain it!) So we can believe for each of the familiar natural numbers (as far as we care to count) that we have implemented it as a set. If I is an inductive set, we can see that (the set implementing) 0 is in I by the definition of "inductive". If the set implementing the familiar natural number n is in I, then (by definition of "inductive") the set implementing the familiar natural number $n+1$ will be in I. So by the principle of mathematical induction, sets implementing each of the familiar natural numbers are in I. But I was *any* inductive set, so for each familiar natural number n, the set implementing n is in the intersection of all inductive sets, that is in $\mathbb{N}$ as we have defined it. This is why we call inductive sets "inductive", by the way. How can we be sure that there aren't some other unintended elements of $\mathbb{N}$? The best argument we can give is

[5]For some purposes, it is useful to modify the definition of $\bigcup A$ so that when x is an atom, $\bigcup\{x\} = x$.

[6]In the system of the unsorted preamble, the discussion above can be adapted to define 0_x as $\{\emptyset_x\}$, an element of $\tau^3(x)$, to define $\sigma(A)$ for any set A as

$$\{a \cup \{x\} \in \tau^{-1}(\tau(A)) : a \in A \wedge x \notin a\},$$

which is of the same type as A, thus allowing definition of $0_x, 1_x, 2_x \ldots$, the natural numbers in $\tau^3(x)$. We can then define $\mathbb{N}_x \in \tau^4(x)$ as the intersection of all inductive sets in $\tau^4(x)$.

this: if there is a collection containing exactly the implementations of the familiar natural numbers, we observe that 0 is certainly in it and $n+1$ must be in it if n is in it. So this collection is inductive, so any element of $\mathbb{N}$, the intersection of all inductive sets, must belong to this set too, and so must be one of the familiar natural numbers. We will see later that there are models of type theory (and of untyped set theory) in which there *are* "unintended" elements of $\mathbb{N}$. In such models the collection of familiar natural numbers must fail to be a set. How can this happen when each type $\mathbf{k+1}$ is supposed to be the collection of *all* sets of type $\mathbf{k}$ objects? Notice that the axiom of comprehension only forces us to implement the subcollections of type $\mathbf{k}$ which are definable using a formula of our language as type $\mathbf{k+1}$ objects. So if there are "unintended" natural numbers we will find that no formula of our language will pick out just the familiar natural numbers. If we insist that each type $\mathbf{k+1}$ contain *all* collections of type $\mathbf{k}$ objects, it will follow that we have defined the set of natural numbers correctly.

Definition: We define $\mathbb{F}$, the set of all finite sets, as $\bigcup \mathbb{N}$. A set which is not finite (not an element of $\mathbb{F}$) is said to be *infinite*.

Since we have defined each natural number n as the set of all sets with n elements, this is the correct definition of finite set (a finite set is a set which has n elements for some natural number n, so exactly a set which belongs to n for some $n \in \mathbb{N}$).

Now we can state a promised axiom.

Axiom of Infinity: $V \notin \mathbb{F}$

This says exactly that the universe is infinite.

In all of this, we have not issued the usual warnings about types. We summarize them here. For $A+1$ to be defined, a set must be of at least type 2. $A+1$ is of the same type as A. Similarly, 0 is of type at least 2 (and there is a formally distinct $0^{\mathbf{n+2}}$ for each n). Any inductive set must be of at least type 3 and the set of all inductive sets $\mathcal{I}$ is of at least type 4. $\mathbb{N}$ is then of type at least 3 (it being the minimal inductive set) and there is actually a $\mathbb{N}^{\mathbf{n+3}}$ in each type $\mathbf{n+3}$. An amusing pun which you may check is $0 \in 1$. The Axiom of Infinity, like the two earlier axioms, says something about each type: the universal set over each type is infinite (it could be written more precisely as $V^{\mathbf{n+1}} \notin \mathbb{F}^{\mathbf{n+2}}$).

We state basic properties of the natural numbers. These are Peano's axioms for arithmetic in their original form. The theory with these axioms (which makes essential use of sets of natural numbers in its formulation) is called *second-order Peano arithmetic*.

1. $0 \in \mathbb{N}$
2. For each $n \in \mathbb{N}$, $\sigma(n) \in \mathbb{N}$.
3. For all $n \in \mathbb{N}$, $\sigma(n) \neq 0$
4. For all $m, n \in \mathbb{N}$, $\sigma(m) = \sigma(n) \rightarrow m = n$.
5. For any set $I \subseteq \mathbb{N}$ such that $0 \in I$ and for all $n \in I$, $\sigma(n) \in I$, all natural numbers belong to I (the principle of mathematical induction).

All of these are obvious from the definition of $\mathbb{N}$ except axiom 4. It is axiom 4 that hinges on the adoption of the Axiom of Infinity.

The principle of mathematical induction (axiom 5) can be presented as another

Proof Strategy: To deduce a goal

$$(\forall n \in \mathbb{N}.\phi[n]),$$

define A as the set $\{n \in \mathbb{N} \mid \phi[n]\}$ and deduce the following goals:

Basis step: $0 \in A$

Induction step: The goal is $(\forall k \in \mathbb{N} \mid k \in A \rightarrow \sigma(k) \in A)$ (or $(\forall k \in \mathbb{N} \mid k \in A \rightarrow k + 1 \in A)$): to prove this, let k be an arbitary natural number, assume $k \in A$ (equivalently $\phi[k]$) (called the *inductive hypothesis*) and deduce the new goal $\sigma(k) \in A$, or $k + 1 \in A$ (equivalently $\phi[\sigma(k)]$, or $\phi[k + 1]$).

We prove some theorems about natural numbers. Our aim is to prove the equivalence of the Axiom of Infinity and Peano's Axiom 4. We will start by trying this and failing, but the nature of our failure will indicate what lemmas we need to prove for ultimate success.

*__Theorem (using Infinity):__ For all $m, n \in \mathbb{N}$, $m + 1 = n + 1 \rightarrow m = n$.

*__Proof:__ Suppose that m and n are natural numbers and $m+1 = n+1$. Our aim is to show that $m = n$. We show this by choosing an arbitrary element a of m and showing that it belongs to n (and also the converse, but this will be direct by symmetry). Suppose $x \notin a$ (we can find such an x by Infinity). Now $a \cup \{x\} \in m+1$, by definition, so it is in $n+1$. It seems that from $a \cup \{x\} \in n+1$, $a \in n$ should follow, but this will require more work. There is certainly no general result that $x \notin a$ and $a \cup \{x\} \in A+1$ implies $a \in A$. Suppose that $A = \{\{0,1\}\}$. Then $\{0,1\} \cup \{2\} = \{0,2\} \cup \{1\} \in A+1$ and $1 \notin \{0,2\}$, but $\{0,2\} \notin A$. We do believe that $a \cup \{x\} \in n+1$ and $x \notin a$ implies $a \in n$, when n is a natural number, but we need to show this.

Theorem (not using Infinity): For any natural number n, if $x \in n+1$ and $y \in x$, then $x - \{y\} \in n$. [an equivalent form is "if $x \cup \{y\} \in n+1$ then $x - \{y\} \in n$"]

Proof: Let $A = \{n \in \mathbb{N} \mid (\forall xy.x \in n+1 \wedge y \in x \rightarrow x - \{y\} \in n)\}$, i.e., the set of all n for which the theorem is true. Our strategy is to show that the set A is inductive. This is sufficient because an inductive set will contain all natural numbers.

First Goal: $0 \in A$

Proof of First Goal: The goal is equivalent to the assertion that if $x \in 0+1$ and $y \in x$, then $x - \{y\} \in 0$. We suppose that $x \in 0+1 = 1$ and $y \in x$: this implies immediately that $x = \{y\}$, whence we can draw the conclusion $x - \{y\} = \{y\} - \{y\} = \emptyset \in 0$, and $x - \{y\} \in 0$ is our first goal.

Second Goal: $(\forall k \in A.k+1 \in A)$

Proof of Second Goal: Let k be an element of A. Assume that $k \in A$: this means that for any $x \in k+1$ and $y \in x$ we have $x - \{y\} \in k$ (this is the inductive hypothesis). Our goal is $k+1 \in A$: we need to show that if $u \in (k+1)+1$ and $v \in u$ we have $u - \{v\} \in k+1$. So we assume $u \in (k+1)+1$ and $v \in u$: our new goal is $u - \{v\} \in k+1$. We know because $u \in (k+1)+1$ that there are $p \in k+1$ and $q \notin p$ such that $p \cup \{q\} = u$. We consider two cases: either $v = q$ or $v \neq q$. If $v = q$ then $u - \{v\} = (p \cup \{q\}) - \{q\} = p$ (because $q \notin p$) and we have $p \in k+1$ so we have $u - \{v\} \in k+1$. In the case where

$v \neq q$, we have $v \in p$, so $p - \{v\} \in k$ by the inductive hypothesis, and $u - \{v\} = (p - \{v\}) \cup \{q\} \in k + 1$ because $p - \{v\} \in k$ and $q \notin p - \{v\}$. In either case we have the desired goal so we are done.

* **Theorem:** If Infinity is false, then Axiom 4 is false.

* **Proof:** If Infinity is false then V is a finite set, so $V \in n$ for some natural number n. We would like to say then that $\{V\} = n$, so $n+1 = \emptyset$ (there is no way to add a new element to V), so $\emptyset \in \mathbb{N}$, and clearly $\emptyset + 1 = \emptyset$, so $\{V\} + 1 = \emptyset + 1 = \emptyset$, but $\{V\} \neq \emptyset$, which gives a counterexample to Axiom 4. This argument is not so much incorrect as incomplete: how do we know that $V \in n$ excludes n having other elements? The following common sense Lemma fixes this: we believe that a finite set with n elements will not have any proper subsets with n elements...

Theorem (not using Infinity): If n is a natural number and $x, y \in n$ and $x \subseteq y$ then $x = y$.

Proof: Let A be the set of natural numbers for which the theorem is true: $A = \{n \in \mathbb{N} \mid (\forall xy.x \in n \wedge y \in n \wedge x \subseteq y \rightarrow x = y)\}$. Our strategy is to show that A is inductive.

First Goal: $0 \in A$

Proof of First Goal: What we need to prove is that if $x \in 0$ and $y \in 0$ and $x \subseteq y$ then $x = y$. Assume that $x \in 0$ and $y \in 0$ and $x \subseteq y$. It follows that $x = \emptyset$ and $y = \emptyset$, so $x = y$. This completes the proof. Note that the hypothesis $x \subseteq y$ did not need to be used.

Second Goal: $(\forall k \in A.k + 1 \in A)$

Proof of Second Goal: Assume $k \in A$. This means that for all $x, y \in k$, if $x \subseteq y$ then $x = y$. This is called the inductive hypothesis.

Our goal is $k + 1 \in A$. This means that for all $u, v \in k + 1$, if $u \subseteq v$ then $u = v$. Suppose that $u \in k + 1$, $v \in k + 1$, and $u \subseteq v$. Our goal is now $u = v$. Because $u \in k + 1$, there are a and b such that $u = a \cup \{b\}$, $a \in k$, and $b \notin a$. Because $u \subseteq v$ we have $a = u - \{b\} \subseteq v - \{b\}$. $a \in k$ has been assumed and $v - \{b\} \in k$

by the previous theorem ($b \in v$ because $u \subseteq v$), so $a = v - \{b\}$ by inductive hypothesis, so $u = a \cup \{b\} = (v - \{b\}) \cup \{b\} = v$.

Theorem (not using Infinity): If there is a natural number n such that $V \in n$, we have $n = \{V\}$, $n + 1 = \emptyset \in \mathbb{N}$, and $n + 1 = \emptyset + 1$, though $n \neq \emptyset$, a counterexample to Axiom 4.

Proof: If $V \in n \in \mathbb{N}$, then for any $x \in n$ we clearly have $x \subseteq V$ whence $x = V$ by the previous theorem, so $n = \{V\}$. That $\{V\} + 1 = \emptyset$ is obvious from the definition of successor (we cannot add a new element to V). It then clearly follows that $\emptyset$ is a natural number. $\emptyset$+1 $= \emptyset$ is also obvious from the definition of successor, so we get the counterexample to Axiom 4.

Theorem (using Infinity): $(\forall mn \in \mathbb{N}.m + 1 = n + 1 \to m = n)$.

Proof: Suppose that m and n are natural numbers and $m + 1 = n + 1$.

We prove that $m = n$ by showing that they have the same elements.

Let $a \in m$ be chosen arbitrarily: our aim is to show $a \in n$.

Choose $x \notin a$ (that there is such an x follows from the Axiom of Infinity, which tells us that the finite set a (finite because it belongs to a natural number) cannot be V). $a \cup \{x\} \in m + 1$. It follows that $a \cup \{x\} \in n + 1$, since by hypothesis $m + 1 = n + 1$. It then follows that $a = (a \cup \{x\}) - \{x\} \in n$ by the first in our sequence of theorems here. This is the goal of the first part of the proof.

In the second part of the proof, we choose $a \in n$ arbitrarily and our goal is to show $a \in m$. The proof is precisely the same as the previous part with m and n interchanged.

So Axiom 4 of Peano arithmetic holds in our implementation.

A familiar construction of finite objects is the construction of *ordered pairs*.

*__ordered pair:__ We define $\langle x, y \rangle$ as $\{\{x\}, \{x, y\}\}$. Note that the pair is two types higher than its components x and y.

Theorem: For any x, y, z, w (all of the same type), $\langle x, y \rangle = \langle z, w \rangle$ iff $x = z$ and $y = w$.

Proof: This is left as an exercise.

*__cartesian product:__ For any sets A and B, we define $A \times B$, the *cartesian product* of A and B, as $\{\langle a, b\rangle \mid a \in A \wedge b \in B\}$. Notice that this is an example of generalized set builder notation, and could also be written as $\{c \mid (\exists ab.c = \langle a, b\rangle \wedge a \in A \wedge b \in B)\}$ (giving a promised example of the generalized set builder notation definition).

The definitions above are starred because we will in fact not use these common definitions. These definitions (due to Kuratowski) are usable in typed set theory and have in fact been used, but they have a practical disadvantage: the pair $\langle x, y\rangle$ is two types higher than its components x and y.

We will instead introduce a new primitive notion and axiom.

ordered pair: For any objects $x^{\mathbf{n}}$ and $y^{\mathbf{n}}$, we introduce primitive notation $\langle x^{\mathbf{n}}, y^{\mathbf{n}}\rangle^{\mathbf{n}}$ for the ordered pair of x and y and primitive notation $\pi_1(x^{\mathbf{n}})^{\mathbf{n}}$ and $\pi_2(x^{\mathbf{n}})^{\mathbf{n}}$ for the first and second projections of an object $x^{\mathbf{n}}$ considered as an ordered pair. As the notation suggests, the type of the pair is the same as the types of its components x and y (which we call its *projections*). In accordance with our usual practice, we will omit the type indices most of the time, allowing them to be deduced from the context.

Notice that the scope of the Axiom of Comprehension is expanded to cover statements including these notations.

Axiom of the Ordered Pair: For any x, y, $\pi_1(\langle x, y\rangle) = x$ and $\pi_2(\langle x, y\rangle) = y$. For any x, $x = \langle \pi_1(x), \pi_2(x)\rangle$.

Corollary: For any x, y, z, w, $\langle x, y\rangle = \langle z, w\rangle \leftrightarrow x = z \wedge y = w$. The corollary is usually taken to be the defining property of the ordered pair; our axiom has the additional consequence that all objects are ordered pairs.

cartesian product: For any sets A and B, we define $A \times B$, the *cartesian product* of A and B, as $\{\langle a, b\rangle \mid a \in A \wedge b \in B\}$. Notice that this is an example of generalized set builder notation, and could also be written as $\{c \mid (\exists ab.c = \langle a, b\rangle \wedge a \in A \wedge b \in B)\}$ (giving a promised example of the generalized set builder notation definition).

We define A^2 as $A \times A$ and more generally define A^{n+1} as $A \times A^n$ (this definition of "cartesian powers" would not work if we were using the Kuratowski pair, for reasons of type). Notice that these exponents can be distinguished from type superscripts (when they are used) because we do not use boldface.

A crucial advantage of a type-level pair in practice is that it allows a nice definition of n-tuples for every n:

tuples: $\langle x_1, x_2, \ldots, x_n\rangle = \langle x_1, \langle x_2, \ldots, x_n\rangle\rangle$ for $n > 2$.

This would not type correctly if the Kuratowski pair were used. We illustrate the problem. If we want to represent $\langle x, y, z\rangle$ as $\langle x, \langle y, z\rangle\rangle$, and assign type n to z, then y will also be assigned type n, $\langle y, z\rangle$ will be assigned type $n + 2$, and x will be assigned type $n + 2$! This can be repaired by using $\langle \iota^{\mathbf{2}}\text{`}x, \langle y, z\rangle\rangle$ instead. The type of the triple thus implemented will be $n + 4$. Now imagine what this approach would give as the definition of a quintuple of objects of the same type: progressively longer tuples defined in this way will be of progressively higher type. We will briefly describe in a later section how the Kuratowski pair can be used to define n-tuples of arbitrary length of the same type independent of n.

We show that the Axiom of Infinity follows from the Axiom of Ordered Pairs (so we strictly speaking do not need the Axiom of Infinity if we assume the Axiom of Ordered Pairs).

Theorem: The Axiom of Ordered Pairs implies the Axiom of Infinity.

Proof: We argue that if $A \in n \in \mathbb{N}$ then $A \times \{0\} \in n$. $\emptyset$ is the only element of 0 and $\emptyset \times \{0\} = 0 \in \mathbb{N}$. Suppose that $A \times \{0\} \in n$ for all $A \in n$. Any element of $n + 1$ is of the form $A \cup \{x\}$ where $A \in n$ and $x \notin A$. $(A \cup \{x\}) \times \{0\} = (A \times \{0\}) \cup \{\langle x, 0\rangle\} \in n + 1$. The claim follows by induction. Now suppose $V \in N \in \mathbb{N}$. It follows that $V \times \{0\} \in N$. But certainly $V \times \{0\} \subseteq V$ so by a theorem about finite sets proved above, $V = V \times \{0\}$, which is absurd.

2.6.1 Digression: The Quine Ordered Pair

We develop a more complex definition of an ordered pair $\langle x, y\rangle$, due to Willard v. O. Quine, which is of the same type as its components x and y and satisfies

the Axiom of Ordered Pairs above, but only works if strong extensionality is assumed.

The definition of the Quine pair is quite elaborate. The basic idea is that the Quine pair $\langle A, B\rangle$ is a kind of tagged union of A and B (it is only defined on sets of sets). Suppose that we can associate with each element a of A an object $\texttt{first}(a)$ from which a can be recovered, and with each element b of B an object $\texttt{second}(b)$ from which b can be recovered, and we can be sure that $\texttt{first}(a)$ and $\texttt{second}(b)$ will be distinct from each other for any $a \in A$ and $b \in B$. The idea is that $\langle A, B\rangle$ will be defined as

$$\{\texttt{first}(a) \mid a \in A\} \cup \{\texttt{second}(b) \mid b \in B\}.$$

For this to work we need the following things to be true for all objects x and y of the type to which elements of A and B belong:

1. For any x, y, $\texttt{first}(x) = \texttt{first}(y) \rightarrow x = y$
2. For any x, y, $\texttt{second}(x) = \texttt{second}(y) \rightarrow x = y$
3. For any x, y, $\texttt{first}(x) \neq \texttt{second}(y)$

If these conditions hold, then we can recover A and B from $\langle A, B\rangle$. An element x of $\langle A, B\rangle$ will be of the form $\texttt{first}(a)$ for some $a \in A$ or of the form $\texttt{second}(b)$ for some $b \in B$. It will be only one of these things, because no $\texttt{first}(x)$ is equal to any $\texttt{second}(y)$. Moreover, if $x = \texttt{first}(a)$, there is only one a for which this is true, and if $x = \texttt{second}(b)$ there is only one b for which this is true. So $A = \{a \mid \texttt{first}(a) \in \langle A, B\rangle\}$ and $B = \{b \mid \texttt{second}(b) \in \langle A, B\rangle\}$.

Thus if $\langle A, B\rangle = \langle C, D\rangle$ we have $A = \{a \mid \texttt{first}(a) \in \langle A, B\rangle\} = \{a \mid \texttt{first}(a) \in \langle C, D\rangle\} = C$ and similarly $B = D$.

The details of the definitions of the needed `first` and `second` operators follow. They will actually be called σ_1 and σ_2.

Definition: For each $n \in \mathbb{N}$ we define $\sigma_0(n)$ as $n + 1$ and for each $x \notin \mathbb{N}$ we define $\sigma_0(x)$ as x. Note that $\sigma_0(x)$ is of the same type as x.

Observation: For any x, y, if $\sigma_0(x) = \sigma_0(y)$ then $x = y$. If x and y are not natural numbers then this is obvious. If x is a natural number and y is not, then $\sigma_0(x)$ is a natural number and $\sigma_0(y)$ is not, so the hypothesis cannot be true. If x and y are natural numbers the statement to be proved is true by axiom 4.

Definition: We define $\sigma_1(x)$ as $\{\sigma_0(y) \mid y \in x\}$. We define $\sigma_2(x)$ as $\sigma_1(x) \cup \{0\}$. We define $\sigma_3(x)$ as $\{y \mid \sigma_0(y) \in x\}$. Note that all of these operations preserve type.

Observation: $\sigma_3(\sigma_1(x)) = x$, so if $\sigma_1(x) = \sigma_1(y)$ we have $x = \sigma_3(\sigma_1(x)) = \sigma_3(\sigma_1(y)) = y$; $\sigma_3(\sigma_2(x)) = x$, so similarly if $\sigma_2(x) = \sigma_2(y)$ we have $x = y$; $\sigma_1(x) \neq \sigma_2(y)$, because $0 \notin \sigma_1(x)$ and $0 \in \sigma_2(y)$. This shows that the σ_1 and σ_2 operations have the correct properties to play the roles of `first` and `second` in the abstract discussion above.

Definition: We define σ_1"(x) as $\{\sigma_1(y) \mid y \in x\}$, σ_2"(x) as $\{\sigma_2(y) \mid y \in x\}$ and σ_3"(x) as $\{\sigma_3(y) \mid y \in x\}$

Definition: We define $\langle x, y\rangle$ as σ_1"$(x) \cup \sigma_2$"(y). Note that the pair is of the same type as its components.

Theorem: For each set x there are unique sets $\pi_1(x)$ and $\pi_2(x)$ such that $\langle \pi_1(x), \pi_2(x)\rangle = x$. An immediate corollary is that for any x, y, z, w (all of the same type), $\langle x, y\rangle = \langle z, w\rangle$ iff $x = z$ and $y = w$.

Proof: $\pi_1(x) = \sigma_3$"$(\{y \in x \mid 0 \notin y\})$; $\pi_2(x) = \sigma_3$"$(\{y \in x \mid 0 \in y\})$

The Quine pair is defined only at type 4 and above; this is not a problem for us because we can do all our mathematical work in as high a type as we need to: notice that the natural numbers we have defined are present in each type above type 2; all mathematical constructions we present will be possible to carry out in any sufficiently high type.

In the theory with weak extensionality, the Quine pair is defined only on sets of sets (elements of $\mathcal{P}^{\mathbf{2}}(V)$) in types 4 and above, but it does satisfy the Axiom of Ordered Pairs on this restricted domain. We could in principle use the Quine pair instead of introducing a primitive pair, if we were willing to restrict relations and functions to domains consisting of sets of sets. This isn't as bad as it seems because all objects of mathematical interest are actually sets of sets.

We will not do this (our primitive pair acts on all objects), but we can use the Quine pair on sets of sets to justify our introduction of the primitive pair: if we cut down our universe to the sets of sets in types 4 and above, and use the relation $x \in' y$ defined as $x \in y \wedge y \in \mathcal{P}^{\mathbf{3}}(V)$ as our new membership relation (allowing only sets of sets of sets to be sets in the restricted world) it is straightforward to verify that our axioms will hold

with the new membership relation and the Quine pair in the old world (with its associated projection functions) will still be a pair and projections in the new world satisfying the Axiom of Ordered Pairs. We can do even better. If we replace the natural numbers n in the definition of the Quine pair in the old world with $n \cap \mathcal{P}^{\mathbf{3}}(V)$, the pair in the new world will turn out to coincide with the new world's Quine pair on sets of sets (because the objects $n \cap \mathcal{P}^{\mathbf{3}}(V)$ are the natural numbers in the new world), and further all pairs of sets will be sets.

We generalize the idea of the previous paragraph. Suppose we have an expression $W_{\mathbf{n}}$ with a type parameter, satisfying $\mathcal{P}(W_{\mathbf{n}}) \subseteq W_{\mathbf{n+1}}$. Notice that if i is the type of elements of $W_{\mathbf{0}}$ then $n+i$ will be the type of elements of $W_{\mathbf{n}}$. Now define $x \in_W y$ as $x \in y \wedge x \in W_{\mathbf{n}} \wedge y \in \mathcal{P}(W_{\mathbf{n}})$ for x of type $n+i$ and y of type $n+i+1$. Define $\mathtt{set}_W(x)$ as $x \in \mathcal{P}(W_{\mathbf{n}})$ for x of type $n+i+1$. Think of $W_{\mathbf{n}}$ as type n of a "W-world" embedded in the world of our type theory (which we will refer to as the "real world" when we need to contrast the worlds). If we have $\mathtt{set}_W(x)$, we have $z \in_W x \leftrightarrow z \in x$ for any z, so if we have $\mathtt{set}_W(x)$ and $\mathtt{set}_W(y)$ and for every z, $z \in_W x$ iff $z \in_W y$, we also have $z \in x$ iff $z \in y$, and of course x and y are sets (since they belong to a power set) so they are equal. We have just shown that extensionality holds in the W-world. Notice that an atom in the sense of the W-world is either an atom in the real world or a set in the real world which has an element which is not in the W-world. Now let $P[x]$ be any sentence of our language. Observe that for any $x \in W_{\mathbf{n}}$ (n being appropriate to the type of x) $P[x] \leftrightarrow x \in \{x \in W_{\mathbf{n}} \mid P[x]\}$, so $P[x] \leftrightarrow x \in_W \{x \in W_{\mathbf{n}} \mid P[x]\}$ (because $x \in W_{\mathbf{n}}$ and $\{x \in W_{\mathbf{n}} \mid P[x]\} \in \mathcal{P}(W_{\mathbf{n}})$), so Comprehension holds in the W-world. Now we note that if we define $W_{\mathbf{n}}$ as $\mathcal{P}^{\mathbf{2}}(V^{\mathbf{n+3}})$, we do have the relation $\mathcal{P}(W_{\mathbf{n}}) \subseteq W_{\mathbf{n+1}}$, as clearly

$$\mathcal{P}(\mathcal{P}^{\mathbf{2}}(V^{\mathbf{n+3}})) = \mathcal{P}^{\mathbf{3}}(V^{\mathbf{n+3}}) \subseteq \mathcal{P}^{\mathbf{2}}(V^{\mathbf{n+4}}).$$

Note further that if $x, y \in W_{\mathbf{n}}$, we also have $\langle x, y\rangle$ (the Quine pair of x and y) belonging to $W_{\mathbf{n}}$, and further $\pi_1(x)$ and $\pi_2(x)$ (the Quine projections of x) belong to $W_{\mathbf{n}}$, so the Axiom of Ordered Pairs is true in the W-world (where the pair is read as the Quine pair of the real world). The definition of $W_{\mathbf{n}}$ is driven by the fact that we need $x, y \in W_{\mathbf{n}}$ to be sets of sets (thus the double power set) and we need the elements of their elements to be of a type which contains natural numbers (thus the double power set of $V^{\mathbf{n+3}}$, the lowest type universal set which contains natural numbers, which are of types $n+2$). A further trick will cause the pair inherited from the larger world

to actually be the Quine pair in the W-world when its projections are sets of sets in the W-world: in the definition of the Quine pair in the real world, replace natural numbers $n^{\mathbf{k+2}} \in \mathbb{N}^{\mathbf{k+3}}$ with their restrictions $n^{\mathbf{k+2}} \cap \mathcal{P}^{\mathbf{3}}(V^{\mathbf{k-1}})$ whenever $k > 0$; this has the effect of replacing the n of the larger world with the n of the W-world, so that the modified Quine pair in the real world is exactly the Quine pair in the W-world when its projections are sets of sets in the W-world. So we can justify the use of the Axiom of Ordered Pairs, if we have the Axiom of Infinity in the real world, by stipulating that we restrict our attention to this W-world henceforth, and we can even preserve the fact that the pair is the Quine pair, though only for sets of sets.

What we have just given is a sketch of what is called a *relative consistency proof*. Given a model of our type theory with the Axiom of Infinity, we show how to get a model of our type theory with the Axiom of Ordered Pairs (but not quite the same model).

Something important is going on here: we are forcibly reminded here that we are *implementing* already familiar mathematical concepts, not revealing what they "really are". Each implementation has advantages and disadvantages. Here, the Kuratowski pair has the advantage of simplicity and independence of use of the Axiom of Infinity, while the Quine pair (or the primitive pair we have to introduce because we allow non-sets) has the technical advantage, which will be seen later to be overwhelming, that it is type level. Neither is the *true* ordered pair; the ordered pair notion prior to implementation is not any particular sort of object: its essence is perhaps expressed in the theorem that equal ordered pairs have equal components. The internal details of the implementation will not matter much in the sequel: what will do the mathematical work is the fact that the pair exactly determines its two components.

2.6.2 Exercises

1. Write a definition of the natural number 2 in the form $\{x \mid \phi[x]\}$ where ϕ is a formula containing only variables, logical symbols, equality and membership. Hint: the formula $\phi[x]$ needs to express the idea that x has exactly two elements in completely logical terms. How would you say that x has at least two elements? How would you say that x has at most two elements?

   A definition of 1 in this style is

   $$\{x \mid (\exists y.y \in x) \wedge (\forall uv.u \in x \wedge v \in x \rightarrow u = v)\}.$$

   Another definition of 1 is

   $$\{x \mid (\exists y.y \in x \wedge (\forall z.z \in x \rightarrow z = y))\}.$$

   Notice the different structure of the scopes of the quantifiers in the two definitions.

2. The usual definition of the ordered pair used in untyped set theory (due to Kuratowski) is

   $$\langle x, y\rangle =_{\mathtt{def}} \{\{x\}, \{x, y\}\}.$$

   We will not use this as our definition of ordered pair because it has the inconvenient feature that the pair is two types higher than its projections. What we *can* do (as an exercise in thinking about sets) is prove the following basic Theorem about this pair definition:

   $$\langle x, y\rangle = \langle z, w\rangle \rightarrow x = z \wedge y = w$$

   This is your exercise. There are various ways to approach it: one often finds it necessary to reason by cases. if you have seen a proof of this, don't go look it up: write your own.

3. Prove the theorem $(\forall xyz.\{x, z\} = \{y, z\} \rightarrow x = y)$ from the axioms of type theory, the definition of unordered pairs $\{u, v\}$, logic and the properties of equality. Remember that distinct letters do not necessarily represent distinct objects.

   This could be used to give a very efficient solution to the previous exercise.

4. Prove that the set $\mathbb{N}^{\mathbf{k+3}}$ (the set of natural numbers in type **k+3**) is inductive. You don't need to specify types on every variable (or constant) every time it occurs, but you might want to state the type of each object mentioned in the proof the first time it appears.

   This proof is among other things an exercise in the careful reading of definitions.

5. Prove the following statement using the Peano axioms in the form stated in the current section: $(\forall n \in \mathbb{N}.n = 0 \vee (\exists m.m + 1 = n))$. You will need to use mathematical induction (in the set based form introduced above), but there is something very odd (indeed rather funny) about this inductive proof.

   Why is the object m unique in case it exists? (This is a throwaway corollary of the main theorem: it does not require an additional induction argument).

6. You are given that $n > 0$ is a natural number and a, b are not natural numbers.

   Compute the Quine pairs $\langle x, y\rangle$ and $\langle y, x\rangle$ where $x = \{\{\emptyset, 3\}, \{2\}, \{0, b\}\}$ and $y = \{\{1, 2\}, \{n, a\}\}$

   Given that $\langle u, v\rangle = \{\{0, 2, 4\}, \{a, b, 2\}, \{0\}, \{1\}, \{a, n\}\}$, what are the sets u and v?

7. Prove that the following are pair definitions (that is, show that they satisfy the defining theorem of ordered pairs).

   The Wiener pair: This is the first ordered pair definition in terms of set theory ever given.

   $$\langle x, y\rangle =_{\mathtt{def}} \{\{\{x\}, \emptyset\}, \{\{y\}\}\}.$$

   Hint: think about how many elements the sets appearing as components of this definition have.

   What is the type of the Wiener pair relative to the types of its projections?

   A pair that raises type by one: This is due to the author. Define $[x, a, b]$ as $\{\{x', a, b\} \mid x' \in x\}$. Define $\langle x, y\rangle$ as $[x, 0, 1] \cup [x, 2, 3] \cup$

$[y, 4, 5] \cup [y, 6, 7]$, where 0,1,2,3,4,5,6,7 can be any eight distinct objects. This only serves to construct pairs of sets, like the Quine pair.

8. We define an *initial segment of the natural numbers* as a set S of natural numbers which has the property that for all natural numbers m, if $m + 1 \in S$ then $m \in S$.

   Does an initial segment of the natural numbers need to contain all natural numbers? Explain why, or why not (with an example).

   Prove that any nonempty initial segment of the natural numbers includes 0.

   How do we prove *anything* about natural numbers?

9. Find sets A and B such that $A + 1 = B + 1$ but $A \neq B$. I found an example that isn't too hard to describe where $A + 1 = B + 1 = 3$ (or any large enough natural number; nothing special about 3). There are other classes of examples. This shows that Axiom 4 is true of natural numbers but not of sets in general.

   Can you describe a set A such that $A + 1 = A$?

10. Verify the equation

$$(A \cup \{x\}) \times \{0\} = (A \times \{0\}) \cup \{\langle x, 0\rangle\}$$

found in the proof that the Axiom of Ordered Pairs implies the Axiom of Infinity. This is an exercise in reading definitions carefully.

We give some solutions.

2. We repeat the definition

$$\langle x, y\rangle =_{\mathtt{def}} \{\{x\}, \{x, y\}\}$$

of the Kuratowski pair. Our goal is to prove

$$(\forall xyzw.\, \langle x, y\rangle = \langle z, w\rangle \to x = z \wedge y = w).$$

We let x,y,z,w be arbitrarily chosen objects. Assume that $\langle x, y\rangle = \langle z, w\rangle$: our new goal is $x = z \wedge y = w$. Unpacking definitions tells us that we have assumed $\{\{x\}, \{x, y\}\} = \{\{z\}, \{z, w\}\}$.

We have two things to prove (since our goal is a conjunction). Note that these are not separate cases: the result proved as the first subgoal can (and will) be used in the proof of the second.

Goal 1: $x = z$

Proof of Goal 1: Because $\{\{x\}, \{x, y\}\} = \{\{z\}, \{z, w\}\}$, we have either $\{x\} = \{z\}$ or $\{x\} = \{z, w\}$. This allows us to set up a proof by cases.

Case 1a: We assume $\{x\} = \{z\}$. Certainly $x \in \{x\}$; thus by substitution $x \in \{z\}$, thus by definition of $\{z\}$ (and by comprehension) we have $x = z$.

Case 1b: We assume $\{x\} = \{z, w\}$. Certainly $z \in \{z, w\}$ (by definition of $\{z, w\}$ and comprehension). Thus $z \in \{x\}$, by substitution of equals for equals. Thus $z = x$, so $x = z$.

Conclusion: In both cases $x = z$ is proved, so Goal 1 is proved.

Goal 2: $y = w$

Proof of Goal 2: Note that we can use the result $x = z$ proved above in this subproof.

Because $\{\{x\}, \{x, y\}\} = \{\{z\}, \{z, w\}\}$ we have either $\{x\} = \{z, w\}$ or $\{x, y\} = \{z, w\}$. This allows us to set up an argument by cases.

Case 2a: Assume $\{x\} = \{z, w\}$. Since $z \in \{z, w\}$ and $w \in \{z, w\}$, we have $z \in \{x\}$ and $w \in \{x\}$ by substitution, whence we have $x = z = w$. This implies that $\{z\} = \{z, w\}$, so

$\{\{z\},\{z,w\}\} = \{\{z\}\}$. Now we have $\{\{x\},\{x,y\}\} = \{\{z\}\}$ by substitution into our original assumption, whence $\{x,y\} = \{z\}$, whence $x = y = z$ (the proofs of these last two statements are exactly parallel to things already proved), so $y = w$ as desired, since we also have $x = z = w$.

Case 2b: Assume $\{x,y\} = \{z,w\}$. Suppose $y \neq w$ for the sake of a contradiction. Since $y \in \{x,y\}$, we have $y \in \{z,w\}$, whence $y = z$ or $y = w$ Since $y \neq w$, we have $y = z$. Since $w \in \{x,y\}$ we have $w = x$ or $w = y$. Since $w \neq y$, we have $w = x$. Now we have $y = z = x = w$, so $y = w$, giving the desired contradiction, and completing the proof that $y = w$.

Conclusion: Since $y = w$ can be deduced in both cases, it can be deduced from our original assumption, completing the proof of Goal 2 and of the entire theorem.

5. Our goal is $(\forall n \in \mathbb{N}.n = 0 \vee (\exists m.m + 1 = n))$.

Define A as the set $\{n \in \mathbb{N} \mid n = 0 \vee (\exists m \in \mathbb{N}.m + 1 = n)\}$.

Our goal is to prove that A is inductive, from which it will follow that $\mathbb{N} \subseteq A$, from which the theorem follows.

Basis Step: $0 \in A$

Proof of Basis Step: $0 \in A \leftrightarrow (0 = 0 \vee (\exists m \in \mathbb{N}.m + 1 = 0))$, and $0 = 0$ is obviously true.

Induction Step: $(\forall k \in \mathbb{N}.k \in A \rightarrow k + 1 \in A)\}$.

Proof of Induction Step: Let k be an arbitrarily chosen natural number. Assume $k \in A$. Our goal is to prove $k + 1 \in A$, that is, $k+1 = 0 \vee (\exists m \in \mathbb{N}.m+1 = k+1)$. We prove this by observing that $k \in \mathbb{N}$ and $k+1 = k+1$, which witnesses $(\exists m \in \mathbb{N}.m+1 = k+1)$. Notice that the inductive hypothesis $k \in A$ was never used at all: there is no need to expand it.

8. We define an *initial segment of the natural numbers* as a set S of natural numbers which has the property that for all natural numbers m, if $m+1 \in S$ then $m \in S$.

Does an initial segment of the natural numbers need to contain all natural numbers? Explain why, or why not (with an example).

Solution: No. The empty set is an initial segment, since the hypothesis $m+1 \in S$ is false for every m if $S = \emptyset$, making $m+1 \in S \rightarrow m \in S$ vacuously true. A nonempty initial segment not equal to $\mathbb{N}$ is for example $\{0,1\}$: the implication can be checked for $m = 0$ and is vacuously true for all other values of m.

Prove that any nonempty initial segment of the natural numbers includes 0.

Solution: Let S be a nonempty initial segment of the natural numbers. Our goal is to show $0 \in S$. Since S is nonempty, we can find $m \in S$. If we could show $(\forall n \in \mathbb{N}.n \in S \rightarrow 0 \in S)$, we would have $m \in S \rightarrow 0 \in S$ and $0 \in S$ by modus ponens.

We prove the lemma $(\forall n \in \mathbb{N}.n \in S \rightarrow 0 \in S)$ by mathematical induction. Let $A = \{n \in \mathbb{N} \mid n \in S \rightarrow 0 \in S\}$. We show that A is inductive.

Basis Step: $0 \in S \rightarrow 0 \in S$ is the goal. This is obvious.

Induction Step: Let k be an arbitrarily chosen natural number. Suppose $k \in A$. Our goal is $k+1 \in A$. $k \in A$ means $k \in A \rightarrow 0 \in S$. We have $k+1 \in S \rightarrow k \in S$ because S is an initial segment. From these two implications $k+1 \in S \rightarrow 0 \in S$ follows, completing the proof of the induction step and the lemma.

2.7 Relations and Functions

If A and B are sets, we define a *relation from A to B* as a subset of $A \times B$. A *relation* in general is simply a set of ordered pairs.

If R is a relation from A to B, we define $x\,R\,y$ as $\langle x, y\rangle \in R$. This notation should be viewed with care. Note here that x and y must be of the same type, while R is one type higher than x or y (that would be three types higher if we used the Kuratowski pair). In the superficially similar notation $x \in y$, y is one type higher than x and $\in$ does not denote a set at all: do not confuse logical relations with set relations. In some cases they can be conflated: the notation $x \subseteq y$ can be used to motivate a definition of $\subseteq$ as a set relation ($[\subseteq] = \{\langle x, y\rangle \mid x \subseteq y\}$), though we do not originally understand $x \subseteq y$ as saying anything about a set of ordered pairs.

If R is a relation, we define $\mathtt{dom}(R)$, the *domain of R*, as $\{x \mid (\exists y.x\,R\,y)\}$. We define R^{-1}, the *inverse of R*, as $\{\langle x, y\rangle \mid y\,R\,x\}$. We define $\mathtt{rng}(R)$, the *range of R*, as $\mathtt{dom}(R^{-1})$. We define $\mathtt{fld}(R)$, the *field of R*, as the union of $\mathtt{dom}(R)$ and $\mathtt{rng}(R)$. If R is a relation from A to B and S is a relation from B to C, we define $R|S$, the *relative product of R and S* as

$$\{\langle x, z\rangle \mid (\exists y.x\,R\,y \wedge y\,S\,z)\}.$$

The symbol $[=]$ is used to denote the equality relation $\{\langle x, x\rangle \mid x \in V\}$. Similarly $[\subseteq]$ can be used as a name for the subset relation (as we did above), and so forth: the brackets convert a grammatical "transitive verb" to a noun.[7]

We define special characteristics of relations.

reflexive: R is *reflexive* iff $x\,R\,x$ for all $x \in \mathtt{fld}(R)$.

symmetric: R is *symmetric* iff for all x and y, $x\,R\,y \leftrightarrow y\,R\,x$.

antisymmetric: R is *antisymmetric* iff for all x, y if $x\,R\,y$ and $y\,R\,x$ then $x = y$.

asymmetric: R is *asymmetric* iff for all x, y if $x\,R\,y$ then $\neg y\,R\,x$. Note that this immediately implies $\neg x\,R\,x$.

transitive: R is *transitive* iff for all x, y, z if $x\,R\,y$ and $y\,R\,z$ then $x\,R\,z$.

[7] The transformation of relation symbols into terms using brackets is an invention of ours and not likely to be found in other books.

equivalence relation: A relation is an *equivalence relation* iff it is reflexive, symmetric, and transitive.

partial order: A relation is a *partial order* iff it is reflexive, antisymmetric, and transitive.

strict partial order: A relation is a *strict partial order* iff it is asymmetric and transitive. Given a partial order R, $R - [=]$ will be a strict partial order. From a strict partial order $R - [=]$, the partial order R can be recovered if it has no "isolated points" (elements of its field related only to themselves).

linear order: A partial order R is a *linear order* iff for any $x, y \in \mathtt{fld}(R)$, either $x\,R\,y$ or $y\,R\,x$. Note that a linear order is precisely determined by the corresponding strict partial order if its domain has two or more elements.

strict linear order: A strict partial order R is a *strict linear order* iff for any $x, y \in \mathtt{fld}(R)$, one has $x\,R\,y$, $y\,R\,x$ or $x = y$. If R is a linear order, $R - [=]$ is a strict linear order.

image: For any set $A \subseteq \mathtt{fld}(R)$, $R\text{“}A = \{b \mid (\exists a \in A.a\,R\,b)\}$.

extensional: A relation R is said to be *extensional* iff for any $x, y \in \mathtt{fld}(R)$, $R^{-1}\text{“}(\{x\}) = R^{-1}\text{“}(\{y\}) \to x = y$: elements of the field of R with the same preimage under R are equal. An extensional relation supports a representation of some of the subsets of its field by the elements of its field.

well-founded: A relation R is *well-founded* iff for each nonempty subset A of $\mathtt{fld}(R)$ there is $a \in A$ such that for no $b \in A$ do we have $b\,R\,a$ (we call this a minimal element of A with respect to R, though note that R is not necessarily an order relation).

well-ordering: A linear order R is a *well-ordering* iff the corresponding strict partial order $R - [=]$ is well-founded.

strict well-ordering: A strict linear order R is a *strict well-ordering* iff it is well-founded.

end extension: A relation S *end extends* a relation R iff $R \subseteq S$ and for any $x \in \mathtt{fld}(R)$, $R^{-1}\text{“}\{x\} = S^{-1}\text{“}\{x\}$. (This is a nonstandard adaptation of a piece of terminology from model theory).

function: f is a *function from A to B* (written $f : A \to B$) iff f is a relation from A to B, $\mathtt{dom}(f) = A$, and for all x, y, z, if $x\, f\, y$ and $x\, f\, z$ then $y = z$. For each $x \in \mathtt{dom}(f)$, we define $f(x)$ as the unique y such that $x\, f\, y$ (this exists because x is in the domain and is unique because f is a function). The notation $f[A]$ is common for the image $f\text{“}A$.

warning about function notation: Notations like $\mathcal{P}(x)$ for the power set of x should not be misconstrued as examples of the function value notation $f(x)$. There is no function $\mathcal{P}$ because $\mathcal{P}(x)$ is one type higher than x. We have considered using the notation $F\text{‘}x$ (this was Russell's original notation for function values) for defined operators in general and restricting the notation $f(x)$ to the case where f is actually a set function. If we did this we would exclude (for example) the notation $\mathcal{P}(x)$ in favor of $\mathcal{P}\text{‘}x$ (or $\mathcal{P}\text{‘}(t)$ for complex terms t that require parentheses). If we used the Russell notation in this way we would also write $\bigcup\text{‘}x, \bigcap\text{‘}x$ because these operations also shift type. We would then prefer the use of $f[A]$ to the use of $f\text{“}A$ for images under functions. But we have not adopted such a convention here.

injection: A function f is an *injection* (or *one-to-one*) iff f^{-1} is a function.

surjection: A function f is a *surjection from A to B* or a *function from A onto B* iff it is a function from A to B and $f\text{“}A = B$.

bijection: A function f is a *bijection from A to B* iff it is an injection and also a surjection from A to B.

composition and restriction: If f is a function and A is a set (usually a subset of $\mathtt{dom}(f)$), define $f\restriction A$ as $f \cap (A \times V)$ (the *restriction of f to the set A*). If f and g are functions and $\mathtt{rng}(g) \subseteq \mathtt{dom}(f)$, define $f \circ g$ as $g|f$. This is called the *composition* of f and g. We may now and then write compositions as relative products, when the unnaturalness of the order of the composition operation is a problem.

identity function: Note that $[=]$ is a function. We call it the *identity function*, and we call $[=]\restriction A$ the *identity function on A*, where A is any set.

abstraction: If $T[x]$ is a term (usually involving x) define $(x : A \mapsto T[x])$ or $(\lambda x : A.T[x])$ as $\{\langle x, T[x]\rangle \mid x \in A\}$. The explicit mention of the set A may be omitted when it is V or when it is understood from the form of the term $T[x]$.

2.7.1 Exercises

1. I give alternative definitions of injection and surjection from A to B.

   A function f is an injection from A to B iff it is a function from A to B and for all $x, y \in A$, $f(x) = f(y) \rightarrow x = y$.

   A function f is a surjection from A to B iff it is a function from A to B and for all $y \in B$, there exists $x \in A$ such that $f(x) = y$.

   Verify that each of these definitions is equivalent to the original one.

2. Prove that if f is an injection from A to B and g is an injection from B to C, then $g \circ f$ is an injection from A to C. ($g \circ f$ may be supposed defined by the equation $(g \circ f)(x) = g(f(x))$)

   Prove that if f is an surjection from A to B and g is a surjection from B to C, then $g \circ f$ is an surjection from A to C.

   Use the alternative definitions of "injection" and "surjection" given in the previous problem and proof strategy as described in chapter 1.

   Comment: of course this shows compositions of bijections are bijections, which will be useful.

3. We outline how to define an n-tuple for arbitrary $n \in \mathbb{N}$ using the Kuratowski pair. Let $\langle x_1, \ldots, x_n\rangle$ be the "function" $\{[i, x_i] \mid i \in \{1, \ldots, n\}\}$, where the notation $[i, x_i]$ is to be read as a Kuratowski pair (in this context we need different notations for the Kuratowski pair and the 2-tuple; explain).

   How do you pick out x_i given $\langle x_1, \ldots, x_n\rangle$ and i?

   What is the relation between the type of the n-tuples and the common type of the x_i's (we do assume that they are all of the same type).

   Give a recursive definition of $\langle x_1, \ldots, x_n\rangle$ in terms of $\langle x_1, \ldots, x_{n-1}\rangle$ and x_n, using explicit set operations. You will need a basis for this recursion (a definition of $\langle\rangle$ or $\langle x\rangle$).

2.8 Digression: The logic of subjects and predicates, or second-order logic

This section is at a higher philosophical level than the preceding. It originally appeared at the end of the Proof chapter, as it is about an extension of our logic, but the level of mathematical sophistication seems to require a prior treatment of ordered pairs and relations, so we have moved it here. In any case, this is not an essential part of our main development.

At the bottom, the subject of logic ought to be completely general: we ought to be able to talk about the entire universe. So we declare that the domain over which the variables x varies in $(\forall x.P[x])$ is simply the domain of all things, whatever things there are.

One might look at a unary sentence $P(x)$ or an atomic sentence $x\,R\,y$ and think that two (respectively three) objects are being discussed: the objects x [resp. x and y] and the predicate P [resp. R].

We are going to analyze this impression. First of all, we simplify matters by reading every unary sentence $P(x)$ as actually having the underlying form $x\,P\,x$, so that all predicates are of the same sort (binary relations). Secondly, we consider the difference between sentences $A[x, y]$ and the atomic predicates R that we are given. If our sentences $A[x, y]$ are meaningful they too must express relations, so we give names $\{x \to y : A[x, y]\}$ for such relations. The rule for using this construction is that $a\{x \to y \mid A[x, y]\}b$ is to mean $A[a, b]$.

We allow predicate variables and quantification over the realm of predicates (= binary relations). For any sentence $\mathbf{P}[R]$ in which a relation symbol R appears, we allow the formation of sentences $(\forall R.\mathbf{P}[R])$ and $(\exists R.\mathbf{P}[R])$. The rules for manipulating these relation quantifiers are exactly the same as for manipulating the quantifiers over objects.

We state firmly that we are *not* admitting a new sort of object (relations) over which these variables range. The objects over which the variables x range are all the objects, and it can be proved that there can be no identification of the relations with a subset of our usual objects.

We add a further abbreviation $\{x \mid R[x]\}$ for $\{x \to y : x = y \wedge R[x]\}$. This ties in with our abbreviation of $R(x)$ as $x\,R\,x$ (the change from brackets to parentheses here is principled!)

Suppose that the relations R are to be identified with some objects. We can preserve the grammatical distinction by writing $\texttt{object}(R)$ for the ob-

ject to be identified with R. Now consider $\mathcal{R}$, a specific relation defined as $\{x \mid (\exists X.x = \texttt{object}(X) \wedge \neg X(x)\}$. $\mathcal{R}(\texttt{object}(\mathcal{R}))$ is then equivalent to $(\exists X.\texttt{object}(\mathcal{R}) = \texttt{object}(X) \wedge \neg X(\texttt{object}(\mathcal{R})))$, which is clearly equivalent to $\neg\mathcal{R}(\texttt{object}(\mathcal{R})))$. This is impossible. So relations in general cannot be objects. This is another analysis of "Russell's paradox".

We stand by our stricture that the domain of our object variables is the entire universe of objects, so we do not allow relation variables to be regarded as denoting objects. Nonetheless, we do not regard it as senseless to say that something is true of all predicates. For example, $R(0) \wedge (\forall x.R(x) \rightarrow R(x+1))$ expresses the idea that a relation R is inductive, and $(\forall R.(\forall R.R(0) \wedge (\forall x.R(x) \rightarrow R(x+1))) \rightarrow R(3))$ is simply a true statement (3 has all inductive properties).

We resist certain extensions of this logical framework (which is usually called "second-order logic").

The first extension we resist is the extension to ternary and higher arity relations. We avoid the necessity to do this by making an assumption about the world:

$$(\exists\Pi_1.(\exists\Pi_2.(\forall xy.(\exists!z.x\Pi_1 z \wedge y\Pi_2 z)) \wedge (\forall z.(\exists!x.x\Pi_1 z) \wedge (\exists!y.y\Pi_2 z))))$$

This asserts the existence of a pairing construction on the universe by asserting the existence of its projection relations. The unique object z such that $x\Pi_1 z$ and $y\Pi_1 z$ whose existence is declared can be called (x, y), the ordered pair of x and y, and a ternary relation $B(x, y, z)$ can be taken as really meaning $x\,B\,(y, z)$ (with similar magic dispelling relations of all higher arities).

The second extension, which is much harder to resist, is the temptation to proceed to logic of third and higher orders.

Formally speaking, to pass to third order logic is to proceed to allow names for objects $\{R \rightarrow S : \mathbf{P}[R, S]\}$ representing binary relations on relations R and S, and then to admit quantifiers over these, and so forth. Other more complex classes of relations and predicates can be imagined.

We can express all consequences of such a move in logic of second order alone. The idea is to specify a domain D_0 to which we restrict the object variables of our original language, then introduce a domain D_1 and a relation E which satisfies $(\forall R.(\exists r \in D_1.(\forall xy \in D_0.x\,R\,y \leftrightarrow (x, y)\,E\,r)))$. The objects in D_0 will be the genuine objects; the objects in D_1 will be (or include) our relations; the true higher order relations will include the relations of third

order alluded to above and indeed all the further strange kinds of relation you might want.

This can be done without enhancing our logic from "mere" second order logic, and moreover, the picture given is false to our intentions. We insist that there are not two tiers of objects in our logic: the domain of the object variables is all objects. So while we can simulate a picture in which there are first order objects, second order objects which capture all relations on first order objects, and third order relations, this is actually not an enhancement of our world, but a (very interesting) suggestion of how there might be a lot of extra complexity in the objects. Note that we can iterate this to obtain fourth order logic, fifth order logic and so forth, and in fact our theory of types above looks remarkably like such an iteration.

We are not tempted in the direction of third, fourth and higher order logic by thinking that the predicates represent a higher tier of objects: we know by the argument above that as we add more and more tiers of relations of various orders the domain of objects we are talking about at the base must depart further and further from being *all* the objects. There is another subtler temptation, which is to introduce third, fourth and higher order logic not as a higher tier of objects but as a higher tier of ... relations. For certainly relations have properties. "R is symmetric" is a perfectly reasonable abbreviation for $(\forall xy.x\,R\,y \leftrightarrow y\,R\,x)$. Our view is that we have not succumbed to the siren lure of third order logic in this direction as long as we only talk about specific properties, relations and operations on predicates. As long as we introduce no variables ranging over predicates of predicates (and woe betide us if we introduce quantifiers over predicates of predicates) we have not advanced to the level of third-order logic.

We give a brief account of what we are doing in the development of our theory of types in terms of the framework of "second-order logic".

We use $x \in U$ to abbreviate $U(x)$ where x is an object and U is a predicate.

We consider the assertion "U is a type" as meaning

$$((\forall xy \in U.(x,y) \in U) \wedge (\exists V.(\exists E.(\forall S.(\exists s \in V.(\forall x \in U.S(x) \leftrightarrow x\,E\,s))))))$$

For U a type and V a predicate, we read

$$(\exists E.(\forall S.(\exists s \in V.(\forall x \in U.S(x) \leftrightarrow x\,E\,s))))$$

as "V is a power domain for U". This says that V contains codes for the restriction of every unary predicate to U (which makes V quite large).

We assert as an axiom that any type has a power domain which is a type.

Type 0 will be a predicate U_0; for each concrete natural number, type U_{i+1} will be a power domain over type U_i with membership relation E_i (which as we will see is not the internal membership relation of the type theory).

$x \in_0 y$, where x is type i and y is type $i+1$, is to be read $x\, E_i\, y$. $x \sim y$ is to be read $x = y$ if x and y are of type 0. Otherwise it is to be read $(\forall z.z \in_0 x \leftrightarrow z \in_0 y)$. $x \in y$ is to be read $x \in_0 y$ if x is of type 0, and otherwise $x \in_0 y \wedge (\forall uv.u \sim v \rightarrow u \in_0 y \leftrightarrow v \in_0 y)$.

The relations $\in$ and $\sim$ in each type can be taken to implement our membership and equality relations for type theory.

A reader should notice that our construction of type theory amounts to iterating the passage to third-order logic which we deprecated, repeatedly. Here we are using this machinery to implement additional complexity in the domain of objects, which we remarked was a sound reason to be interested in this kind of structure.

It should be noted that our type theory is an entirely "first-order" theory: there are no quantifiers over predicates in its language. As a result, it may have interesting "first-order" models in which not all restrictions of predicates to the types define sets at higher types, and we will see much later that this is the case. The idea is that the comprehension axiom asserts that all sets of type i objects defined by statements in the language of type theory exist; this is not the same as saying (as we do in the framework presented here) that in some sense all sets of type i objects are implemented at type $i+1$.

On a very technical level, it should be noted that there is no way whatsoever to define a full infinite sequence of types using the framework we have given here: this does nonetheless support the validity of all reasoning in type theory, because any particular argument in type theory mentions only finitely many types.

2.9 Defining Functions by Recursion; First-Order Peano Arithmetic

Recursion is a special technique for defining functions with domain $\mathbb{N}$.

Informally, a recursive definition might look like this (this is not a completely general example): $f(0) = 0$; for each natural number n, $f(n+1) = (f(n)+1)+1$. This seems somehow suspect because this definition of f appears to mention f itself in an essential way.

We show that this kind of definition is legitimate. We begin by exhibiting the technique of *iterative* definition of which the example just given is a special case.

Iteration Theorem: For any function $f : D \to D$ and $a \in D$ (of appropriate types) there is a unique function $g : \mathbb{N} \to D$ such that $g(0) = a$ and $g(n+1) = f(g(n))$ for each $n \in \mathbb{N}$.

Definition: Where a, f, g are as in the statement of the Theorem, we define $f^n(a)$ as $g(n)$.

Proof of Iteration Theorem: We begin with a nonce

> **Definition:** A set I is said to be (f,a)-*inductive* iff $\langle 0, a\rangle \in I$ and $(\forall nx.\, \langle n, x\rangle \in I \to \langle n+1, f(x)\rangle \in I)$.
>
> Let g be the intersection of all (f,a)-inductive sets. We claim that g is the desired function. Note that we do not even know that g is a function at this point!
>
> We claim that g is a subset of $\mathbb{N} \times D$. Note that $\langle 0, a\rangle \in \mathbb{N} \times D$ and for any $\langle n, x\rangle \in \mathbb{N} \times D$ we also have $\langle n+1, f(x)\rangle \in \mathbb{N} \times D$, so $\mathbb{N} \times D$ is (f,a)-inductive, whence $g \subseteq \mathbb{N} \times D$.
>
> So we now know that every element of g is an ordered pair whose first component is a natural number and whose second component is in D, which is necessary but not sufficient for g to be a function with domain the set of natural numbers and range included in D.
>
> We claim that for each natural number n there is exactly one object x such that $\langle n, x\rangle$ is an element of g. Define A as the set of all natural numbers n such that there is exactly one object x such that $\langle n, x\rangle$ is an element of g: we prove our claim by showing that A is inductive.

We first need to show that $0 \in A$. We know that $\langle 0, a\rangle \in g$, so there is at least one x such that $\langle 0, x\rangle \in g$. Now consider $g' = g - \{\langle 0, x\rangle \mid x \neq a\}$. We claim that g' is (f, a)-inductive. $\langle 0, a\rangle \in g'$ is obvious. Suppose $\langle n, x\rangle \in g'$. It follows that $\langle n+1, f(x)\rangle \in g$, and in fact that $\langle n+1, f(x)\rangle \in g'$, because $\langle n+1, f(x)\rangle \notin \{\langle 0, x\rangle \mid x \neq a\}$. Since g' is (f, a)-inductive, $g \subseteq g'$. But $g' \subseteq g$ as well, so $g = g'$, and a is the only object such that $\langle 0, a\rangle \in g' = g$, which is what we needed to show.

Now we need to show that for any $k \in A$ we also have $k+1 \in A$. Assume $k \in A$, whence there is exactly one u such that $\langle k, u\rangle \in g$. We need to show that there is exactly one v such that $\langle k+1, v\rangle \in g$. Since $\langle k, u\rangle \in g$, it follows that $\langle k+1, f(u)\rangle \in g$, so there is at least one such v. Now define g' as $g - \{\langle k+1, w\rangle \mid w \neq f(u)\}$. We claim that g' is (f, a)-inductive. Clearly $\langle 0, a\rangle \in g'$. Suppose $\langle n, x\rangle \in g'$; our aim is to show $\langle n+1, f(x)\rangle \in g'$. Suppose otherwise for the sake of a contradiction. Clearly $\langle n+1, f(x)\rangle \in g$: it is thus necessary that $\langle n+1, f(x)\rangle \in \{\langle k+1, w\rangle \mid w \neq f(u)\}$, which implies $f(x) \neq f(u)$ and also that $n+1 = k+1$. From this it follows that $n = k$, and thus, since $\langle n, x\rangle = \langle k, x\rangle \in g$, that $x = u$, whence $f(x) \neq f(u)$ is impossible, which is the desired contradiction. We then have $g = g'$, whence $f(u)$ is the only object x such that $\langle k+1, x\rangle \in g' = g$, whence $k+1 \in A$.

This completes the proof that g is a function from $\mathbb{N}$ to D. Since $\langle 0, a\rangle \in g$, we have $g(0) = a$. Since $\langle n, g(n)\rangle \in g$, we have $\langle n+1, f(g(n))\rangle \in g$, whence $g(n+1) = f(g(n))$.

Now we need to show that g is the unique function with these properties. Suppose $g' : \mathbb{N} \to V$, $g'(0) = a$ and $g'(n+1) = f(g'(n))$. $\langle 0, a\rangle \in g'$ is immediate. If $\langle n, x\rangle \in g'$, then $x = g'(n)$, and $\langle n+1, g'(n+1)\rangle = \langle n+1, f(g'(n))\rangle = \langle n+1, f(x)\rangle \in g'$, so g' is (f, a)-inductive, whence $g \subseteq g'$. g' contains exactly one element with first projection n for each natural number n, which must be the one element with first projection n belonging to g, so g and g' are the same set.

This completes the proof of the Iteration Theorem.

Observation: This is more than a technical theorem: it has some philosophically interesting content. Our definition of the natural numbers is based intellectually on the use of natural numbers to count the elements of sets. Here we are showing that our logical machinery allows

us to implement the arguably quite different basic idea of applying an operation n times to an object.

Recursion Theorem: For any set a and function $g : (\mathbb{N} \times V) \to V$, there is a function $h : \mathbb{N} \to V$ such that $h(0) = a$ and $h(n+1) = g(n, h(n))$ for each $n \in \mathbb{N}$.

Proof of Recursion Theorem: Let $G(\langle n, x\rangle)$ be defined as $\langle n+1, g(n,x)\rangle$. Then $h(n) = \pi_2(G^n(\langle 0, a\rangle))$.

There is an alternative way to define $f^n(a)$.

Definition: A set S of natural numbers is an *initial segment of the natural numbers* iff for all $n \in \mathbb{N}$, $n+1 \in S \to n \in S$.

Theorem: Any nonempty initial segment of the natural numbers contains 0.

Theorem: $y = f^n(a)$ iff there is a function g such that the domain of g is an initial segment S of the natural numbers including n as an element, $g(0) = a$, for all m such that $m+1 \in S$ we have $g(m+1) = f(g(m))$, and $y = g(n)$. This formulation is advantageous because it only appeals to the existence of finite sets.

We summarize the basic properties of $f^n(x)$ in the

Recursive definition of iteration: Where D is a set, $a \in D$ and $f : D \to D$,

1. $f^0(a) = a$
2. $f^{n+1}(a) = f(f^n(a))$

The types of a and n in $f^n(a)$ are the same, and the type of f is one higher. Notice that type indices are bold-faced, so they will not be confused with these indices.

We also define the freestanding notation f^n for the function

$$\{\langle x, f^n(x)\rangle : x \in D\}.$$

When we iterate an operation which is not a function, as for example in $\mathcal{P}^{\mathbf{2}}(A)$, the power set of the power set of A, we bold-face the index (which in this case cannot possibly be a type index) to indicate that this is not an example of iteration. There is in fact no reference to the number 2 (of any type) in this expression at all, any more than there is a reference to 2 when we write a variable $x^{\mathbf{2}}$ with the type index 2.

As examples we can present definitions of addition and multiplication.

Give the nonce name σ to the successor function on natural numbers ($\sigma = \{(n, n+1) \mid n \in \mathbb{N}\}$) We can define $m+n$ (for any natural numbers m, n) as $\sigma^n(m)$ (adding n is iterating successor n times). We can define $m \cdot n$ for any natural numbers m and n as $(\sigma^m)^n(0)$: to add $m \cdot n$ is to add m n times.

At this point we can observe that our original definition of $n+1$ and the new definition of $n+1$ in terms of the addition function just defined agree. If n is a natural number, $n+1$ (read as a sum) is defined as $\sigma^1(n)$ which is $\sigma^{\sigma(0)}(n)$ by the definition of 1, which is $\sigma(\sigma^0(n))$ by an application of the definition of $f^n(a)$, which is $\sigma(n)$ by another application of the definition of $f^n(a)$, which, finally, is $n+1$ in the original sense (the successor of n) by definition of the function σ.

Demonstrations of some properties of addition and multiplication using the recursive definition of iteration:

1. $m + 0 = \sigma^0(m) = m$, so $m + 0 = m$.
2. $m + \sigma(n) = \sigma^{n+1}(m) = \sigma(\sigma^n(m)) = \sigma(m+n)$, so $m + \sigma(n) = \sigma(m+n)$, or $m + (n+1) = (m+n) + 1$.
3. $m \cdot 0 = (\sigma^m)^0(0) = 0$, so $m \cdot 0 = 0$
4. $m \cdot \sigma(n) = (\sigma^m)^{n+1}(0) = \sigma^m((\sigma^m)^n(0)) = \sigma^m(m \cdot n) = m \cdot n + m$, so $m \cdot \sigma(n) = m \cdot n + m$, or $m \cdot (n+1) = m \cdot n + m$.

The recursive (really as we see above "iterative") definitions of addition and multiplication are incorporated into modern formulations of "Peano's axioms", which make no essential reference to sets. The theory with these axioms is formally called *first-order Peano arithmetic*.

When we reason in first-order Peano arithmetic, we are not reasoning in our type theory. But, since we have shown that there is an intepretation of the axioms of first-order Peano arithmetic in our type theory, any theorems

we prove in first-order Peano arithmetic will be true in that interpretation. We will see below that there is a different interpretation of Peano arithmetic commonly used in untyped set theory (the von Neumann definition of the natural numbers, already mentioned above), and anything we prove in arithmetic will also be true in that interpretation (and in any other we come up with).

The convention when reasoning in first-order Peano arithmetic is to assume that all quantifiers are restricted to the natural numbers (we are not talking about anything else, and notably we are not talking about sets of natural numbers as we do in the original (second-order) version of the theory). Note this particularly in axiom 5.

1. 0 is a natural number.
2. For each natural number n, $\sigma(n)$ is a natural number. For all natural numbers m, n, $m + n$ and $m \cdot n$ are natural numbers.
3. For all natural numbers n, $\sigma(n) \neq 0$
4. For all natural numbers m, n, $\sigma(m) = \sigma(n) \rightarrow m = n$.
5. For each formula $\phi[n]$, we adopt as an axiom $\phi[0] \wedge (\forall k.\phi[k] \rightarrow \phi[\sigma(k)]) \rightarrow (\forall n.\phi[n])$. This is the principle of mathematical induction. Note that this is not really a single axiom 5, but a suite of axioms 5_ϕ. Such a suite is called an *axiom scheme*. A scheme is needed because we do not refer to sets here.
6. For all natural numbers m, n, $m + 0 = m$
7. For all natural numbers m, n, $m + \sigma(n) = \sigma(m + n)$
8. For all natural numbers m, n, $m \cdot 0 = 0$
9. For all natural numbers m, n, $m \cdot \sigma(n) = m \cdot n + m$

Since addition is also a primitive operation here, we use a primitive notation for successor at first rather than the more natural addition of 1. Notice the reformulation of mathematical induction in terms of formulas rather than sets. This formulation of mathematical induction is not a statement with a quantifier over formulas (we cannot really do that for reasons which we may discuss *much* later on) but an infinite collection of different axioms, one for

each formula ϕ. You should notice that the axioms for addition and multiplication capture the iterative definitions of addition and multiplication given above.

We give some sample proofs in Peano arithmetic.

Definition: $1 = \sigma(0)$ (of course this recapitulates an earlier definition given in the context of our type theory). Note that it is immediate from the axioms for addition that $n + 1 = n + \sigma(0) = \sigma(n + 0) = \sigma(n)$. We feel free to use these notations interchangeably.

Proof Strategy: We give the first order version of mathematical induction as a proof strategy.

To deduce a goal $(\forall n.\phi[n])$, deduce the following two goals:

Basis step: Deduce $\phi[0]$.

Induction step: Deduce $(\forall k.\phi[k] \rightarrow \phi[k + 1])$. Application of prior proof strategy expands this: let k be an arbitrarily chosen natural number (which might be 0!): assume $\phi[k]$ (this is called the *inductive hypothesis*, and it is useful to emphasize where in an induction proof the inductive hypothesis is used), and deduce the new goal $\phi[k + 1]$.

Theorem: For each natural number $n \neq 0$, there is a unique natural number m such that $m + 1 = n$.

Proof: We prove by mathematical induction the assertion "For each natural number n, if $n \neq 0$, then there is a natural number m such that $m+1 = n$".

For $n = 0$ this is trivially true (basis step).

Suppose it is true for $n = k$; then our goal is to prove that it is true for $n = k + 1$ (induction step).

Either $k = 0$ or there is an m such that $m + 1 = k$, by inductive hypothesis. In either case, there is an m' such that $m' + 1 = k + 1$, namely k itself.

So the assertion is true for all n by mathematical induction. What is strange here is that the inductive hypothesis is not used in this proof!

The observant reader will notice that we have not yet proved the theorem. We have shown that for each nonzero natural number n there is an m such that $m + 1 = n$, but we have not shown that this m is unique yet. Suppose that $m + 1 = n$ and also $m' + 1 = n$: it follows directly from axiom 4 that $m = m'$. So we have shown that there can only be one such m for each n and the proof is complete.

Theorem: For each natural number n, $0 + n = n + 0$.

Proof: We prove this by mathematical induction.

$0 + 0 = 0 + 0$ completes the proof of the basis step.

Now for the induction step. We assume that $0+k = k+0$ and our goal is to show that $0 + \sigma(k) = \sigma(k) + 0$. $0 + \sigma(k) = \sigma(0 + k)$ by axioms, and $\sigma(0 + k) = \sigma(k + 0)$ (by inductive hypothesis) $= \sigma(k) = \sigma(k) + 0$. This completes the proof of the induction step and of the theorem.

Theorem: For any natural numbers m, n, $(m + 1) + n = (m + n) + 1$.

We fix m and prove this by induction on n.

The basis step is established by $(m + 1) + 0 = m + 1 = (m + 0) + 1$.

The hypothesis of the induction step is $(m + 1) + k = (m + k) + 1$; the goal is to show $(m+1)+(k+1) = (m+(k+1))+1$. $(m+1)+(k+1) = ((m + 1) + k) + 1$ by axiom, which is equal to $((m + k) + 1) + 1$ by inductive hypothesis, which is in turn equal to $(m + (k + 1)) + 1$ by axiom, completing the proof.

Theorem: For any natural numbers m, n, $m + n = n + m$.

Proof: We prove this by (you guessed it!) mathematical induction.

The statement we actually prove by mathematical induction is "for any natural number n, for any natural number m, $m + n = n + m$."

The basis step is "For any natural number m, $m + 0 = 0 + m$". We just proved that!

The induction hypothesis is "For any natural number m, $m+k = k+m$" (for some fixed natural number k) and the induction goal is "For any natural number m, $m + (k + 1) = (k + 1) + m$". Now $m + (k + 1) = (m+k)+1$ by axiom, which is in turn equal to $(k+m)+1$ by inductive

hypothesis, which is equal to $(k + 1) + m$ by the previous theorem, proving the induction goal and completing the proof of the theorem.

We follow our own example in earlier sections on logic and recapitulate a proof of the same theorem (the commutativity of addition) in a more formal style.

Our aim is to prove $(\forall yx.x+y = y+x)$, the commutative law of addition.

We must be proving it by math induction, as we have no other way to do it!

We will prove this by induction on y (as a rule, it is better to do induction on the variable farthest to the right in an expression you are going to work with, because of the forms of axioms 6-9).

The basis step will be $(\forall x.x + 0 = 0 + x)$

The induction hypothesis will be $(\forall x.x+k = k+x)$ (k being an arbitrary number we introduce).

The induction goal will be $(\forall x.x + \sigma(k) = \sigma(k) + x)$

This gives us the following proof outline:

Goal: $(\forall yx.x + y = y + x)$ We prove this by induction on y.

Basis Goal 1: $(\forall x.x + 0 = 0 + x)$

Let k be chosen arbitrarily.

Ind Hyp 1: $(\forall x.x + k = k + x)$

Induction Goal: $(\forall x.x + \sigma(k) = \sigma(k) + x)$

We now proceed to fill in the complete proof (though not without further comments about what we are doing!!!) The reason we are numbering the basis and induction items is that there will be subproofs of this proof which are induction proofs themselves and have their own bases and induction steps.

Goal: $(\forall yx.x + y = y + x)$ We prove this by induction on y.

Basis Goal 1: $(\forall x.x + 0 = 0 + x)$ We prove this by induction on x!

Basis Goal 2: $0 + 0 = 0 + 0$

1: $0 + 0 = 0 + 0$ ref = [That was easy!]

Ind Hyp 2 (2): $k + 0 = 0 + k$

Ind Goal 2: $\sigma(k) + 0 = 0 + \sigma(k)$

3: $0 + \sigma(k) = \sigma(0 + k)$ ax 7 $x := 0; y := k$

4: $0 + \sigma(k) = \sigma(k + 0)$ subs using (2) [the ind hyp] into (3)

5: $k + 0 = k$ ax 6 $x := k$

6: $0 + \sigma(k) = \sigma(k)$ subs using line (5) into line (4)

7: $\sigma(k) = \sigma(k) + 0$ ax 6 $x := \sigma(k)$

8: $0 + \sigma(k) = \sigma(k) + 0$ trans = 6,7

9: $\sigma(k) + 0 = 0 + \sigma(k)$ symm = 8, and we are done with the basis goal. I proved this differently than I did in class (I think) though the basic idea is the same.

Let k be chosen arbitrarily.

Ind Hyp 1 (2): $(\forall x.x + k = k + x)$ This is line 2 again because everything in an induction step uses local hypotheses and goes away. We could even call it line 1, since we are never going to refer to line 1 again, but the original line 1 has not vanished. It wouldn't do any harm to call this line 10, as long as you know that lines 2-9 above can't be used.

Induction Goal: $(\forall x.x + \sigma(k) = \sigma(k) + x)$ I'm going to start working on the left side of this because I can see what to do with it. I will get as far as I can and then I will see something else that I want to prove...by induction of course.

Let m be arbitrary (I'm not going to use l because it looks too much like a 1). Notice that I'm using the standard technique to deal with a universal quantifier instead of induction. Sometimes it works!

Goal: $m + \sigma(k) = \sigma(k) + m$

3: $m + \sigma(k) = \sigma(m + k)$ ax 7 $x := m, y := k$ working on left side as I said, because axiom 7 applies.

4: $m+\sigma(k) = \sigma(k+m)$ subs using (2) (the ind hyp) into (3).

Goal: I see that I need to prove $\sigma(k) + m = \sigma(k + m)$ to complete the proof. I prove this as a Lemma below, by induction: actually the Lemma is

$$(\forall xy.\sigma(x) + y = \sigma(x + y)),$$

a statement which looks rather like axiom 7 but isn't. The proof of the Lemma is given below; we proceed with the main proof assuming that we have it.

5: $\sigma(k)+m = \sigma(k+m)$ Lemma proved below, $x := k, y := m$

6: $m + \sigma(k) = \sigma(k + m)$ symm trans = lines 4 and 5

This completes the proof of the main theorem, once we prove the Lemma, whose free-standing proof follows. Of course we cannot use commutativity of addition in the proof of the Lemma!

Lemma: $(\forall xy.\sigma(x) + y = \sigma(x + y))$ Im going to prove this by induction on y; first I'm going to use the usual strategy for a universal quantifier to get rid of x.

Let a be chosen arbitrarily.

Goal: $(\forall y.\sigma(a)+y = \sigma(a+y))$ *This* is what we will prove by induction on y.

Basis Goal: $\sigma(a) + 0 = \sigma(a + 0)$

1: $\sigma(a) + 0 = \sigma(a)$ ax 6 $x := \sigma(a)$

2: $a + 0 = a$ ax 6 $x := a$

3: $\sigma(a + 0) = \sigma(a)$ both sides line 2

4: $\sigma(a) + 0 = \sigma(a + 0)$ symm trans = 1,3

Let k be arbitrary

Ind Hyp (5): $\sigma(a) + k = \sigma(a + k)$

Ind Goal: $\sigma(a)+\sigma(k) = \sigma(a+\sigma(k))$ Notice that both sides offer opportunities to calculate using axiom 7.

6: $\sigma(a) + \sigma(k) = \sigma(\sigma(a) + k)$ ax 7, $x := \sigma(a), y := k$ You should notice the opportunity to rewrite using the inductive hypothesis!

7: $\sigma(a)+\sigma(k) = \sigma(\sigma(a+k))$ subs into line 6 using line 5 (the ind hyp)

8: $a+\sigma(k) = \sigma(a+k)$ ax 7 $x := a, y := k$

9: $\sigma(a)+\sigma(k) = \sigma(a+\sigma(k))$ subs using (8) into (7) [an equation can be used to substitute in either order].

That completes the proof of the lemma and the theorem.

Much more natural definitions of the arithmetic operations which use the intuitive idea that the numbers are sizes of sets are given below, and in terms of these definitions much more natural proofs of properties such as the ones just proved can be given. Proofs in Peano arithmetic are nonetheless a useful exercise: they apply to quite different implementations of the natural numbers (another implementation will be given later): for any implementation, if the Peano axioms hold, then all the theorems following from the Peano axioms also hold.

Apparently stronger forms of both induction and recursion are available, but turn out to be equivalent to the basic forms already given. A presentation of these requires some prior discussion of the familiar order on the natural numbers.

Definition: For natural numbers m, n, we say $m \leq n$ (*m is less than or equal to n*) just in case $(\exists k.m + k = n)$. We define $m < n$ (*m is less than n*) as $m \leq n \wedge m \neq n$. We define $m \geq n$ (*m is greater than or equal to n*) as $n \leq m$, and similarly define $m > n$ (*m is greater than n*) as $n < m$.

Note that we assume here that such things as the associative and commutative laws of addition have already been proved.

Theorem: For all natural numbers m, n, k, if $m + k = n + k$ then $m = n$.

Proof: Fix m and n and prove by induction on k. This is obvious for $k = 0$. If it is true for k and $m+(k+1) = n+(k+1)$, then $(m+k)+1 = (n+k)+1$ by addition axiom, $m + k = n + k$ by axiom 4, and $m = n$ by inductive hypothesis.

Theorem: The relation $\leq$ on natural numbers just defined is a linear order.

Proof: $n \leq n = n + 0$ is immediate. If $m \leq n$ and $n \leq m$ then we have $n = m+k$ and $m = n+l$ for some k and l, whence $n = n+0 = n+(k+l)$, so $k + l = 0$, whence it is easy to show that $k = l = 0$, so $m = n$. If $m \leq n$ and $n \leq p$, then for some k, l, $m + k = n$ and $n + l = p$, so $(m + k) + l = m + (k + l) = p$. This shows that $\leq$ is a partial order.

We need to show further that $(\forall m.(\forall n.m \leq n \vee n \leq m))$. That $(\forall n.0 \leq n \wedge n \leq 0)$ is evident, because $0 \leq n$ is true for any n. Suppose that $(\forall n.k \leq n \vee n \leq k)$. We want to show that $(\forall n.k + 1 \leq n \vee n \leq k + 1)$ Either n is 0, in which case $n \leq k + 1$, or for some m, $n = m + 1$, in which case $k \leq m \leftrightarrow m \leq k$ by inductive hypothesis, whence $k + 1 \leq m + 1 = n \vee n = m + 1 \leq k + 1$ by axiom 4.

Theorem: $m \leq n \leftrightarrow m + k \leq n + k$.

$m + p = n \leftrightarrow (m + k) + p = n + k$

Corollary: $m < n \leftrightarrow m + k < n + k$

Theorem: For all $n \in \mathbb{N}$, for all $k \in \mathbb{N}$, $k \leq n \leftrightarrow k < n + 1$.

Proof: Prove this by induction on n. The basis step requires us to show that $m \leq 0 \leftrightarrow m < 1$ for all m. If $m \leq 0$, then since $0 \leq 1$ and $1 \not\leq 0$, $m < 1$ is obvious. If $m \neq 0$ then $m = n + 1$ for some n, so $1 \leq m$, thus $m < 1 \rightarrow m = 0$ (by contrapositive). Now if $m \leq k \leftrightarrow m < k + 1$, for all m, we immediately have $(m + 1) \leq (k + 1) \leftrightarrow (m + 1) < (k + 1) + 1$ We certainly also have $0 \leq k + 1 \leftrightarrow 0 < (k + 1) + 1$, and since every number is either 0 or a successor we have shown for all m that $m \leq k + 1 \leftrightarrow m < (k + 1) + 1$

Theorem (Strong Induction, set form): For any set A of natural numbers, if $(\forall a \in \mathbb{N}.(\forall x < a.x \in A) \rightarrow a \in A)$, then $A = \mathbb{N}$.

Proof: Suppose that A is a set of natural numbers and $(\forall a \in \mathbb{N}.(\forall x < a.x \in A) \rightarrow a \in A)$. We define the set B as $\{b \in \mathbb{N}.(\forall x \leq b.x \in A)\}$. We show that B is inductive. Since $B \subseteq A$ is obvious, $B = \mathbb{N} \rightarrow A = \mathbb{N}$.

Since $(\forall x < 0.x \in A)$ is vacuously true, $0 \in A$. For any $b \leq 0$, $b = 0 \in A$, so $0 \in B$.

Now suppose that $k \in B$. Our goal is to show that $k + 1 \in B$. Since $k \in B$, we have $p \in A$ for all $p \leq k$, and so for all $p < k + 1$. It then

follows that $k+1 \in A$, and since we have $p \in A$ for all $p < k+1$ as well, we also have $k+1 \in B$. This completes the proof that B is inductive, which we have already seen is sufficient for the proof of the theorem.

Theorem (Strong Induction, property form): For any formula ϕ, $(\forall a \in \mathbb{N}.(\forall x < a.\phi[x]) \rightarrow \phi[a]) \rightarrow (\forall n \in \mathbb{N}.\phi[n])$.

Proof: This is proved in the same way as the previous theorem.

There is a form of recursion which is to standard recursion (or iteration) roughly as strong induction is to standard induction.

Theorem (Course-of-Values Recursion): Let A be a set. Let $\mathcal{F}$ be the set of all functions with domain a proper initial segment

$$\{m \in \mathbb{N} \mid m < n\}$$

of the natural numbers and range a subset of A (notice that the function with domain $\emptyset$ is one of these: set $n = 0$). Let G be any function from $\mathcal{F}$ to ι"A. Then there is a uniquely determined function $f : \mathbb{N} \rightarrow A$ such that $\{f(n)\} = G(f\lceil\{m \in \mathbb{N} \mid m < n\})$ for each $n \in \mathbb{N}$.

Proof: We define a function H from $\mathcal{F}$ to $\mathcal{F}$ as follows. If $g \in \mathcal{F}$ has domain $\{m \in \mathbb{N} \mid m < n\}$, define $H(g)$ as $g \cup (\{n\} \times G(g))$ (recall that $G(n)$ is the singleton set containing the intended value at n of the function being constructed). Now apply the iteration theorem: define $f(n)$ as $H^{n+1}(\emptyset)(n)$. It is straightforward to verify that this function has the desired property.

Example: An example of a function defined in this way, in which the value of f at any natural number depends on its values at *all* smaller natural numbers, would be $f(n) = 1 + \Sigma_{i<n} f(i)$[8]

It is a usual exercise in a book of this kind to prove theorems of Peano arithmetic up to the point where it is obvious that the basic computational axioms of arithmetic and algebra can be founded on this basis (and we may do all of this in these notes or in exercises). It is less obvious that all usual notions of arithmetic and algebra can actually be defined in terms of the

[8] I should give more examples of this kind of function definition and a discussion of the interesting things going on here with types.

quite restricted vocabulary of Peano arithmetic and logic: this is very often asserted but seldom actually demonstrated. We supply an outline of how this can be established.

We give basic definitions without (or with only an indication of) supporting proofs to indicate that the expressive power of Peano arithmetic without set language is enough to talk about finite sets of natural numbers and to define recursive functions. This is a serious question because the definition of recursive functions above relies strongly on the use of sets. Notice that we use the alternative formulation of the definition of $f^n(a)$ in this development, because we only code finite sets of natural numbers as natural numbers here, and the alternative formulation has the advantage that it only talks about finite sets.

Definition: For natural numbers m, n we say $m|n$ (n *is divisible by* m or m *is a factor of* n) iff there is a natural number x such that $m \cdot x = n$.

Definition: A natural number p is a *prime* iff it has exactly two factors. (One of these factors must be 1 and the other $p \neq 1$ itself).

Definition: Let p be a prime. A natural number q is a *power of* p iff p is a factor of every factor of q except 1.

Definition: Let p be a prime and n a natural number. A nonzero natural number m *occurs in the base* p *expansion of* n just in case n can be expressed in the form $a \cdot q + m \cdot r + s$, where $q > r > s$ and q, r are powers of p.

The underlying idea is that we now have the ability to code finite sets of natural numbers as natural numbers (and so in fact sets of sets, sets of sets of sets, and so forth).

Definition: Define $x \in_p y$ as "$x + 1$ occurs in the base p expansion of y". For any prime p and naturals $x_1, \ldots, x_n$ all less than $p - 1$ define $\{x_1, \ldots, x_n\}_p$ as the smallest natural number y such that $(\forall z. z \in_p y \leftrightarrow z = x_1 \vee \ldots \vee z = x_n)$. [there is something to prove here, namely that there is such a y].

Definition: Define $\langle x, y \rangle_{p,q}$ as $\{\{x\}_p, \{x, y\}_p\}_q$.

Definition: For any function f, we say that f is *definable in Peano arithmetic* iff there is a formula $\phi[x, y]$ in the language of arithmetic such that $\phi[x, y] \leftrightarrow y = f(x)$.

Theorem: For any function f definable in Peano arithmetic, $y = f^n(x)$ iff there are primes $p < q < r$ such that there is a natural number g such that $(\forall m \leq n.(\exists! y.\, \langle m, y\rangle_{p,q} \in_r g))$ and $\langle 0, x\rangle \in_r g$ and $(\forall m < n.(\forall y.\, \langle m, y\rangle_{p,q} \in_r g \rightarrow \langle m+1, f(y)\rangle_{p,q} \in_r g))$. Note that this is expressible in the language of Peano arithmetic, so all functions definable by iteration of definable functions are definable (and functions definable by recursion from definable functions are also definable since we can represent pairs of natural numbers as natural numbers and define the projection functions of these pairs).

Definition: Define $d(x)$ as $2 \cdot x$. Define 2^n as $d^n(1)$. Define $x \in_{\mathbb{N}} a$ as

$$(\exists y > x.(\exists z < 2^x.(\exists u.a = u \cdot 2^y + 2^x + z))).$$

This expresses that the nth digit in the binary expansion of a is 1, and this supports a nice coding of finite sets of natural numbers as natural numbers, which we will have occasion to use later.

2.9.1 Exercises

1. If I define a function I_n such that $I_n(f) = f^n$ (so for example $I_3(f)(x) = f^3(x) = f(f(f(x)))$, I invite you to consider the functions $(I_n)^m$. For example, compute $(I_2)^3(f)(x)$. Compute $(I_3)^2(f)(x)$. There is an equation $(I_m)^n = I_{F(m,n)}$, where F is a quite familiar operation on natural numbers, which you can write and might derive if you do enough experiments. There is a serious formal problem with this equation, though, in our type theory. What is the function $F(m, n)$? What is the formal problem?

2. Prove the theorem
$$(\forall m : m \neq m + 1)$$
of Peano arithmetic.

   Indicate each application of an axiom and of an inductive hypothesis. Do not apply theorems you have not proved yourself on your paper. You may identify $\sigma(x)$ and $x + 1$ without comment for any natural number x.

3. Prove as many of the following as you can in first-order Peano arithmetic, not necessarily in the given order (but this is the suggested order). Your proofs should not mention sets or the type theory definitions of the natural numbers (this is all just arithmetic from the Peano axioms).

   Use proof strategy. You can be a little more freeform than heretofore, but take pains to make it clear what you are doing. You may use theorems already proved in the notes or already proved by you. You may *not* use anything else you think you know about arithmetic.

   In some of these, you may need to prove lemmas as I had to in the proof of the commutative law of addition.

   I suggest looking at the proof of the left distributive law which appears after these exercises as a style model.

   Do prove at least two of them.

   (a) The associative law of addition.

(b) The distributive law of multiplication over addition (for this one, I require you to prove the right distributive property (a proof of the left distributive property appears below) without using the commutative law of multiplication.

(c) The associative law of multiplication.

(d) The commutative law of multiplication.

4. Prove the *Well-Ordering Principle*: for any nonempty set of natural numbers A, there is an element m of A such that for all $x \in A$, $m \leq x$. The usual hint: how do we prove anything about natural numbers?

5. Assuming ordinary knowledge about elementary algebra of natural numbers and basic properties of divisibility, write a proof by strong induction that any natural number is a finite product of primes.

   As a footnote to this, see if you can make a proposal as to how to formally define the notion of a product of a finite list of primes in our formal system.

2.9.2 A case study: proof of the left distributive law in formal arithmetic

In this subsection, we prove the left distributive law $a \cdot (b + c) = a \cdot b + a \cdot c$. We assume that the associative law of addition

$$(\forall m : (\forall n : (\forall r : (m + n) + r = m + (n + r))))$$

has already been proved (it appears in the previous set of exercises. This proof might be used as a style manual for the formal arithmetic proofs in that exercise set.

Theorem: $(\forall x : \forall y : \forall z : x \cdot (y+z) = x \cdot y + x \cdot z$. I do assume standard order of operations for addition and multiplication without further comment.

(1): $a = a$ trivial line for universal generalization

(2): $b = b$ trivial line for universal generalization

Goal: $(\forall z : a \cdot (b + z) = a \cdot b + a \cdot z)$ (I'm not going to indent here: you need to keep track of the block structure of the proof.)

We prove this goal by mathematical induction.

Basis Goal: $a \cdot (b + 0) = a \cdot b + a \cdot 0$

style remark: We introduce all instances of axioms at the top level (applying UI to the axiom) and then substitute using these instances into other equations. We may often start with a trivial equation, an instance of the reflexivity of equality, as here.

(3): $a \cdot (b + 0) = a \cdot (b + 0)$ refl = (an acceptable abbreviation for "reflexivity of equality")

(4): $b + 0 = b$ UI ax 6 m:=b

(5): $a \cdot (b + 0) = a \cdot b$ subs 4 into 3 (apply substitution property of equality using equation 4 to equation 3)

(6): $a \cdot b = a \cdot b + 0$ UI ax 6 m:=b

(7): $a \cdot (b+0) = a \cdot b+0$ trans = 5,6 (this could also be justified by substitution)

(8): $a \cdot 0 = 0$ UI ax 8 m:= a

(9): $a \cdot (b + 0) = a \cdot b + a \cdot 0$ subs 8 into 7 (substitution in either order can be justified this way: the reader can presumably tell which way we are going).

This completes the proof of the basis goal.

Induction Step:

Ind Hyp: Let k be arbitrarily chosen. Assume **(10)** $a \cdot (b + k) = a \cdot b + a \cdot k$

Ind Goal: $a \cdot (b + \sigma(k)) = a \cdot b + a \cdot \sigma(k)$

(11): $a \cdot (b + \sigma(k)) = a \cdot (b + \sigma(k))$ refl =

(12): $b + \sigma(k) = \sigma(b + k)$ UI ax 7 m:=b n:=k

(13): $a \cdot (b + \sigma(k)) = a \cdot (\sigma(b + k))$ subs 12 into 11

(14): $a \cdot (\sigma(b + k)) = a \cdot (b + k) + a$ UI ax 9 m:= a n:=b+k

(14a): $a \cdot (b + \sigma(k)) = a \cdot (b + k) + a$ trans = 13,14 (accidentally left out a line and didn'f want to renumber the rest of the proof!)

(15): $a \cdot (b + \sigma(k)) = (a \cdot b + a \cdot k) + a$ subs 10 (ind hyp!) into 14a

You **must** supply parentheses. Though we do have associativity of addition by hypothesis, we are still at a level where we show its use explicitly.

(16): $(a \cdot b + a \cdot k) + a = a \cdot b + (a \cdot k + a)$ assoc + [m:= $a \cdot b$ n:= $a \cdot k$ r:=a]

(17): $a \cdot (b + \sigma(k)) = a \cdot b + (a \cdot k + a)$ trans = 15,16

(18): $a \cdot \sigma(k) = a \cdot k + a$ ax 9 m:= a n:= k

(19): $a \cdot (b + \sigma(k)) = a \cdot b + a \cdot \sigma(k)$ subs 18 into 17

(20): $(\forall z : a \cdot (b + z) = a \cdot b + a \cdot z)$ mathematical induction 9, 10-19. Notice that the basis step reference is to a line but the induction step reference must be to a block.

(21): The theorem UG 1-20 (really this is two applications of UG but I think you follow).

In this proof I followed a much more careful strategy than I followed in class, which I suggest for the proof exercises. I started by setting the left side of the theorem equal to itself, then did a series of substitutions into the right side of this equation until I got the desired right side. This might be familiar to you, as it is often mandated as a style of proving trig identities.

Please be aware that you cannot prove equations the way you learned to solve them in algebra. If we have an equation $a = b$ and do the same thing to both sides of it (getting something like $F[a] = F[b]$) we can then posit $F[a] = F[b]$ (this derived rule appears in some proofs above, labelled "both side"), but proving $F[a] = F[b]$ will not prove $a = b$ (this style of reasoning, very useful when trying to solve equations, is not reversible).

2.10 Equivalence Relations, Partitions, and Representatives: the Axiom of Choice

Definition: Sets A and B are said to be *disjoint* just in case $A \cap B = \emptyset$.

Definition: A collection P of sets is said to be *pairwise disjoint* just in case

$$(\forall A \in P.(\forall B \in P.A = B \vee A \cap B = \emptyset)).$$

Definition: A collection P of sets is a *partition of* A iff $\emptyset \notin P$, $\bigcup P = A$, and P is pairwise disjoint. A partition of A is a collection of nonempty sets which do not overlap and which cover all of A. We say that a collection P *is a partition* iff it is a partition of $\bigcup P$.

Definition: If R is an equivalence relation and $x \in \mathtt{fld}(R)$ we define $[x]_R$, the *equivalence class of* x *under* R, as $R\text{“}(\{x\}) = \{y \mid x\,R\,y\}$.

Theorem: If R is an equivalence relation, $P_R = \{[x]_R \mid x \in \mathtt{fld}(R)\}$ is a partition of $\mathtt{fld}(R)$.

Proof: Let R be an arbitrarily chosen equivalence relation. Define $P_R = \{[x]_R \mid x \in \mathtt{fld}(R)\}$.

Our goal is to prove that P_R is a partition of $\mathtt{fld}(R)$. Using the definition of partition, this reduces to three subgoals.

Goal 1: $\emptyset \notin P_R$. Suppose for the sake of a contradiction that $\emptyset \in P_R$. By the definition of P_R as a complex set abstract, this is equivalent to the assertion that $\emptyset = [x]_R$ for some $x \in \mathtt{fld}(R)$. Choose such an x. $x\,R\,x$ holds because R is reflexive, whence $x \in [x]_R$ by the definition of equivalence class, whence $x \in \emptyset$, which yields the desired contradiction. This completes the proof of Goal 1.

Goal 2: $\bigcup P_R = \mathtt{fld}(R)$. Use the proof strategy for showing the equality of two sets.

2a: Let x be an arbitrarily chosen element of $\bigcup P_R$: our new goal is to show $x \in \mathtt{fld}(R)$. Since $x \in \bigcup P_R$, we can choose a set A such that $x \in A$ and $A \in P_R$. Since $A \in P_R$, we can choose y such that $A = [y]_R$. $x \in A = [y]_R$ implies immediately that $y\,R\,x$, whence $x \in \mathtt{fld}(R)$, which completes the proof of goal 2a.

2b: Let x be an arbitrarily chosen element of $\texttt{fld}(R)$: our new goal is to show that $x \in \bigcup P_R$. Since $x \in \texttt{fld}(R)$, we may choose a y such that one of $x\,R\,y$ or $y\,R\,x$ is true. But then both are true because R is symmetric, and we have $x \in [y]_R$. From $x \in [y]_R$ and $[y]_R \in P_R$, we deduce $x \in \bigcup P_R$, completing the proof of goal 2b.

Since any element of either set has been shown to belong to the other, the two sets are equal, completing the proof of Goal 2.

Goal 3: P_R is pairwise disjoint. Our goal is to show that for any elements A, B of P_R we have $A = B \vee A \cap B = \emptyset$ To prove this, we assume that A and B are distinct and take $A \cap B = \emptyset$ as our new goal. We prove this by contradiction: assume $A \cap B \neq \emptyset$ and our new goal is a contradiction. Since $A \cap B \neq \emptyset$, we may choose an $x \in A \cap B$. Since $A, B \in P_R$ we may choose y and z such that $A = [y]_R$ and $B = [z]_R$. If we had $y = z$ we would have $A = B$ and a contradiction, so we must have $y \neq z$. $x \in A \cap B = [y]_R \cap [z]_R$ implies $x \in [y]_R$ and $x \in [z]_R$, whence we have $x\,R\,y$ and $x\,R\,z$, whence by symmetry and transitivity of R we have $y\,R\,z$. We now prove $A = [y]_R = [z]_R = B$, which will give the desired contradiction since A and B were initially supposed distinct.

3a: Let u be an arbitrarily chosen element of $[y]_R$. Our new goal is $u \in [z]_R$. $u \in [y]_R$ implies $y\,R\,u$, and $y\,R\,z$ and symmetry imply $z\,R\,y$. Thus by transitivity of R we have $z\,R\,u$ and so $u \in [z]_R$. This completes the proof of goal 3a.

3b: Let u be an arbitrarily chosen element of $[z]_R$. Our new goal is $u \in [y]_R$. $u \in [z]_R$ implies $z\,R\,u$, which in combination with $y\,R\,z$ and transitivity of R implies $y\,R\,u$, which implies $u \in [y]_R$, which completes the proof of goal 3b.

Since the sets $[y]_R = A$ and $[z]_R = B$ have the same elements, it follows that they are equal, which completes the proof of a contradiction, from which Goal 3 and the Theorem follow.

Theorem: If $\mathcal{P}$ is a partition of A, the relation

$$\equiv_{\mathcal{P}} = \{\langle x, y\rangle \mid (\exists B \in \mathcal{P}.x \in B \wedge y \in B)\}$$

is an equivalence relation with field A.

Proof: This is left as an exercise.

Observation: Further, $\equiv_{P_R} = R$ and $P_{\equiv_{\mathcal{P}}} = \mathcal{P}$ for any R and $\mathcal{P}$: there is a precise correspondence between equivalence relations and partitions.

An equivalence relation R represents a way in which elements of its field are similar: in some mathematical constructions we wish to *identify* objects which are similar in the way indicated by R. One way to do this is to replace references to an $x \in \mathtt{fld}(R)$ with references to its equivalence class $[x]_R$. Note that for all x, y in $\mathtt{fld}(R)$ we have $x\,R\,y$ iff $[x]_R = [y]_R$.

It might be found inconvenient that $[x]_R$ is one type higher than x. In such a situation, we would like to work with a representative of each equivalence class.

Definition: Let P be a partition. A *choice set* for P is a set C with the property that $B \cap C$ has exactly one element for each $B \in P$.

A choice set for the partition P_R will give us exactly one element of each equivalence class under R, which we can then use to represent all elements of the equivalence class in a context in which R-equivalent objects are to be identified.

In some situations, there is a natural way to choose an element of each equivalence class (a canonical representative of the class). We will see examples of this situation. In the general situation, we can invoke the last axiom of our typed theory of sets.

Axiom of Choice: If P is a partition (a pairwise disjoint set of nonempty sets) then there is a choice set C for P.[9]

The Axiom of Choice is a somewhat controversial assertion with profound consequences in set theory: this seemed like a good place to slip it in quietly without attracting too much attention.

Here we also add some terminology about partial orders.

It is conventional when working with a particular partial order $\leq$ to use $<$ to denote $[\leq] - [=]$ (the corresponding strict partial order), $\geq$ to denote

[9]If we include the Hilbert symbol in our logic and allow its use in comprehension, this axiom is not needed: a choice set for P is definable as $(\{(\epsilon x : x \in A) \mid A \in P\}$. This would be a way to slip the Axiom of Choice in while attracting even less attention.

$[\leq]^{-1}$ (which is also a partial order) and $>$ to denote the strict partial order $[\geq] - [=]$.

A minimum of $\leq$ is an element m of $\mathtt{fld}(\leq)$ such that $m \leq x$ for all $x \in \mathtt{fld}(x)$. A maximum of $\leq$ is a minimum of $\geq$. A minimal element with respect to $\leq$ is an element m such that for no x is $x < m$. A maximal element with respect to $\leq$ is a minimal element with respect to $\geq$. Notice that a maximum or minimum is always unique if it exists. A minimum is always a minimal element. The converse is true for linear orders but not for partial orders in general.

For any partial order $\leq$ and $x \in \mathtt{fld}(\leq)$, we define $\mathtt{seg}_{\leq}(x)$ as $\{y \mid y < x\}$ (notice the use of the strict partial order) and $(\leq)_x$ as $[\leq] \cap (\mathtt{seg}_{\leq}(x))^2$. The first set is called the *segment* in $\leq$ determined by x and the second is called the *segment restriction* determined by x.

For any subset A of $\mathtt{fld}(\leq)$, we say that an element x of $\mathtt{fld}(\leq)$ is a lower bound for A in $\leq$ iff $x \leq a$ for all $a \in A$, and an upper bound for A in $\leq$ iff $a \leq x$ for all $a \in A$. If there is a lower bound x of A such that for every lower bound y of A, $y \leq x$, we call this the greatest lower bound of A, written $\inf_{\leq}(A)$, and if there is an upper bound x of A such that for all upper bounds y or A, we have $x \leq y$, we call this the least upper bound of A, written $\sup_{\leq}(A)$.

A special kind of partial order is a *tree*: a partial order $\leq_T$ with field T is a *tree* iff for each $x \in T$ the restriction of $\leq_T$ to $\mathtt{seg}_{\leq_T}(x)$ is a well-ordering. A subset of T which is maximal in the inclusion order among those well-ordered by $\leq_T$ is called a *branch*.

2.10.1 Exercises

1. Suppose that P is a partition.

   Prove that the relation $\sim_P$ defined by
   $x \sim_P y$ iff $(\exists A \in P. x \in A \wedge y \in A)$
   is an equivalence relation. What is the field of this equivalence relation?

   Describe its equivalence classes.

   This is an exercise in carefully writing everything down, so show all details of definitions and proof strategy, as far as you can.

2. This question relies on ordinary knowledge about the reals and the rationals, and also knowledge of Lebesgue measure if you have studied this (if you haven't, don't worry about that part of the question).

   Verify that the relation on real numbers defined by "$x\,R\,y$ iff $x - y$ is rational" is an equivalence relation. It might be cleaner to consider the relation "$x\,R\,y$ iff $x - y$ has a terminating decimal expansion"; the result is similar and the equivalence classes are easier to describe.

   Describe the equivalence classes under this relation in general. Describe two or three specific ones. Note that each of the equivalence classes is countably infinite (why? [see the next section if you don't know what "countably infinite" means]), distinct equivalence classes are disjoint from each other, and so we "ought" to be able to choose a single element from each class.

   Can you think of a way to do this (you will not be able to find one, but thinking about why it is difficult is good for you)?

   Suppose we had a set X containing exactly one element from each equivalence class under R. For each rational number q, let X_q be the set $\{r + q \mid r \in X\}$. Note that X_q is just a translation of X.

   Prove that $\{X_q \mid q \in \mathbb{Q}\}$ is a partition of $\mathbb{R}$. (This will include a proof that the union of the X_q's is the entire real line).

   If you know anything about Lebesgue measure, you might be able to prove at this point that X is not Lebesgue measurable (if you can, do so). It is useful to note that the collection of X_q's is countable.

2.11 Cardinal Number and Arithmetic

We say that two sets are the same size iff there is a one-to-one correspondence (a bijection) between them.

Definition: We say that sets A and B are *equinumerous* and write $A \sim B$ just in case there is a bijection f from A onto B.

Theorem: Equinumerousness is an equivalence relation.

Indication of Proof: It is reflexive because the identity function on any set is a function. It is symmetric because the inverse of a bijection is a bijection. It is transitive because the composition of two bijections is a bijection.

Definition: For any set A, we define $|A|$, the *cardinality of* A, as $[A]_\sim = \{B \mid B \sim A\}$. Notice that $|A|$ is one type higher than A. We define `Card`, the set of all *cardinal numbers*, as $\{|A| \mid A \in V\}$.

The same definitions would work if we were using the Kuratowski pair, and in fact the cardinals would be precisely the same sets.

We have already encountered some cardinal numbers.

Theorem: Each natural number is a cardinal number.

Proof: $|\emptyset| = \{\emptyset\} = 0$ is obvious: there is a bijection from $\emptyset$ to A iff $A = \emptyset$.

Suppose that $n \in \mathbb{N}$ is a cardinal number: show that $n+1$ is a cardinal number and we have completed the proof that all natural numbers are cardinals by mathematical induction. Let x be an element of n. There is a $y \notin x$ because $x \neq V$ (by the Axiom of Infinity). It suffices to show $n + 1 = |x \cup \{y\}|$. To show this, we need to show that for any set z, $z \in n+1$ iff $z \sim x \cup \{y\}$. If $z \in n+1$ then $z = v \cup \{w\}$ for some $v \in n$ and some $w \notin v$. Because n is a cardinal number there is a bijection f from x to v: $f \cup \{\langle y, w\rangle\}$ is readily seen to still be a bijection. Now let z be an arbitrarily chosen set such that $z \sim x \cup \{y\}$. This is witnessed by a bijection f. Now $f^{-1}\text{“}x$ belongs to n because n is a cardinal number, and thus we see that $v = f^{-1}\text{“}x \cup \{f^{-1}(y)\}$ belongs to $n+1$ (certainly $f^{-1}(y) \notin f^{-1}\text{“}x$), completing the proof.

There is at least one cardinal number which is not a natural number.

Definition: We define $\aleph_0$ as $|\mathbb{N}|$. Sets of this cardinality are said to be *countable* or *countably infinite*. Infinite sets not of this cardinality (if there are any) are said to be *uncountable* or *uncountably infinite*.

We provide some lemmas for construction of bijections from other bijections.

Lemma: The union of two relations is of course a relation. The union of two functions is a function iff the functions agree on the intersection of their domains: that is, if f and g are functions, $f \cup g$ is a function iff for every $x \in \mathtt{dom}(f) \cap \mathtt{dom}(g)$ we have $f(x) = g(x)$, or, equivalently but more succinctly, $f\lceil \mathtt{dom}(f) \cap \mathtt{dom}(g) = g\lceil \mathtt{dom}(f) \cap \mathtt{dom}(g)$. Note that it is sufficient for the domains of f and g to be disjoint.

Definition: A function f is said to *cohere* with a function g iff $f\lceil(\mathtt{dom}(f) \cap \mathtt{dom}(g)) = g\lceil(\mathtt{dom}(f) \cap \mathtt{dom}(g))$.

Lemma: The union of two injective functions f and g is an injective function iff f coheres with g and f^{-1} coheres with g^{-1}. Note that it is sufficient for the domain of f to be disjoint from the domain of g and the range of f disjoint from the range of g.

Lemma: For any x and y, $\{\langle x, y\rangle\}$ is an injection.

Arithmetic operations have natural definitions.

A cardinal $|A|$ is the collection of all sets of the same size as A. Thus, if κ is a cardinal, we mean by "set of size κ" simply an element of κ. This is not true of all representations of cardinality: if we used a representative set the same size as A as $|A|$, for example, then a set of size κ would be a set equinumerous with κ (the representation used in the usual set theory introduced later is of this latter kind).

We define addition of cardinals. Informally, a set of size $\kappa + \lambda$ will be the union of two disjoint sets, one of size κ and one of size λ.

Definition (abstract definition of addition): If κ and λ are cardinals, we define $\kappa + \lambda$ as

$$\{A \cup B \mid A \in \kappa \wedge B \in \lambda \wedge A \cap B = \emptyset\}.$$

Notice that this agrees (for cardinal numbers) with the abstract definition of addition of arbitrary sets given in the alternative definition in the initial definition of natural numbers.

There are some things to verify about this definition. One has to verify that $\kappa + \lambda$ is nonempty. If $A \in \kappa$ and $B \in \lambda$ then $A \times \{0\} \in \kappa$, $B \times \{1\} \in \lambda$, and these sets are obviously disjoint. The fact that cartesian product is a type level operation is crucial here (so Infinity is required). One has to verify that $\kappa + \lambda$ is a cardinal.

Observation: $|A| + |B| = |(A \times \{0\}) \cup (B \times \{1\})|$.

Proof: Suppose A' and B' are disjoint sets with bijections $f : A \to A'$ and $g : B \to B'$. Then $(\pi_1|f) \cup (\pi_1|g)$ is a bijection from $(A \times \{0\}) \cup (B \times \{1\})$ to $A' \cup B'$. The union of these two injections is an injection because they have disjoint domains and disjoint ranges, and the union has the correct domain and range.

It is perhaps preferable to simply take the Observation as the

Definition (concrete definition of addition): $|A| + |B|$ is defined as

$$|(A \times \{0\}) \cup (B \times \{1\})|.$$

(It is straightforward to show that this does not depend on the choice of representatives A and B from the cardinals).

The abstract definition of addition would work if we were using Kuratowski pairs but the proof that addition is total would be somewhat harder. The Observation would be incorrect and in fact would not make sense because it would not be well-typed.

Notice that the definition of $\kappa + 1$ as an addition of cardinals agrees with the definition of $\kappa + 1$ as a set already given in the development of finite number.

Before discussing multiplication, we consider the notion of being the same size appropriate to sets at different types.

Definition (alternative notation for singleton set): We recall that we defined $\iota(x)$ as $\{x\}$. The point of this notation is that it is iterable: we can use $\iota^{\mathbf{n}}(x)$ to denote the n-fold singleton of x. [But do notice that

this is not an example of iteration as ι is not a function (a function does not raise type). The n in $\iota^{\mathbf{n}}(x)$ is a purely formal bit of notation (like a type index) and not a reference to any natural number in our theory, and this is why it is in boldface]

Definition (singleton image operations): We define $\iota^{\mathbf{n}}$"x, the *n-fold singleton image of x* as $\{\iota^{\mathbf{n}}(y) \mid y \in x\}$. For any relation R, we define $R^{\iota^{\mathbf{n}}}$ as $\{\langle \iota^{\mathbf{n}}(x), \iota^{\mathbf{n}}(y)\rangle \mid x\, R\, y\}$. We define $T^{\mathbf{n}}(\kappa)$ for any cardinal κ as $|\iota^{\mathbf{n}}\text{"}A|$ for any $A \in \kappa$. Note that $A \sim B \leftrightarrow \iota\text{"}A \sim \iota\text{"}B$ is obvious: if f is a bijection from A to B, then f^{ι} will be a bijection from ι"A to ι"B. We define $T^{-\mathbf{n}}(\kappa)$ as the unique cardinal λ (if there is one) such that $T^{\mathbf{n}}(\lambda) = \kappa$.

Definition (sole element): We define $\iota^{-\mathbf{1}}(\{x\})$ as x. We define $\iota^{-\mathbf{1}}(A)$ as $\emptyset$ if A is not a singleton. $\iota^{-\mathbf{n}}(\iota^{\mathbf{n}}(x))$ will be defined as x as one might expect, if this notation is ever needed.

The singleton map (or iterated singleton map) is in a suitable external sense injective, so a set equinumerous with $\iota^{\mathbf{n}}$"A, though it is n types higher than A, is in a recognizable sense the same size as A.

The definition of $T^{-\mathbf{n}}$ depends on the observation that $T^{\mathbf{n}}$ is "injective" in the sense that $T^n(\mu) = T^n(\nu) \to \mu = \nu$ for any cardinals μ, ν (the scare quotes are needed because T^n cannot actually be viewed as an injection, since it is not a function at all, due to its effect on types) , so if there is a suitable λ there is only one. We leave open the possibility that $T^{-\mathbf{n}}(\kappa)$ is undefined for some cardinals κ and indeed this turns out to be the case.

We discuss the application of the T operation to natural numbers.

Theorem: $T(0) = 0$ and $T(n+1) = T(n) + 1$.

Corollary: $T(1) = 1; T(2) = 2; T(3) = 3 \ldots$ But we cannot say

$$(\forall n \in \mathbb{N}. T(n) = n),$$

because this is ungrammatical.

Theorem: For all natural numbers n, $T(n)$ is a natural number. For all natural numbers n [not of the lowest possible type] $T^{-\mathbf{1}}(n)$ exists and is a natural number.

Proof: We prove both parts by induction, of course.

Our first goal is to prove that $T(n)$ is a natural number for every natural number n. We observe first that $T(0) = 0$ is obvious, as $\iota``\emptyset = \emptyset$. Now suppose that $k$ is a natural number and $T(k)$ is a natural number. Our aim is to prove that $T(k+1)$ is a natural number. Each element of $k+1$ is of the form $A \cup \{x\}$ where $A \in k$ and $x \notin A$. $T(k+1) = |\iota``(A \cup \{x\})|$. But $\iota``(A \cup \{x\}) = \iota``A \cup \{\{x\}\}$. Obviously $\iota``A \in T(k)$ and $\{x\} \notin \iota``A$, so $\iota``A \cup \{\{x\}\} \in T(k) + 1 \in \mathbb{N}$, so $T(k+1) = T(k) + 1 \in \mathbb{N}$.

Our second goal is to prove that $T^{-\mathbf{1}}(n)$ exists and is a natural number for each natural number n (not of the lowest possible type). Since $T(0) = 0$, we also have $T^{-\mathbf{1}}(0) = 0$, so $T^{-\mathbf{1}}(0)$ exists and is a natural number. Let k be a natural number such that there is a natural number l such that $T(l) = k$ (which is equivalent to saying that $T^{-\mathbf{1}}(k)$ exists and is a natural number). Choose a set A of cardinality l. Choose $x \notin A$. $|A \cup \{x\}| = l + 1$ and $|\iota``(A \cup \{x\})| = |\iota``A \cup \{\{x\}\}| = k + 1$ is obvious, so $T(l+1) = k+1$, whence $T^{-\mathbf{1}}(k+1)$ exists and is a natural number as desired.

Reasonable Convention: It is reasonable to simply identify the natural numbers at different types and there is a way to make sense of this in our notation: allow a natural number variable n of type k to appear at other types with the understanding that where it appears in a position appropriate for a variable of type $k + i$ it is actually to be read as $T^i(n)$. We will not do this, or at least we will explicitly note use of this convention if we do use it, but it is useful to note that it is possible.

Rosser's Counting Theorem: $\{1, \ldots, n\} \in T^{\mathbf{2}}(n)$, for each positive natural number n.

Discussion and Proof: Of course $\{1, \ldots, n\} = \{m \in \mathbb{N} \mid 1 \leq m \leq n\}$ has n members, if n is a concrete natural number. But the second n we mention is two types higher than the first one: we fix this by affixing $T^{\mathbf{2}}$ to the second one, so that both occurrences of n have the same type.

What this actually says is that if we have a set A belonging to a natural number n, we can put $\iota^{\mathbf{2}}``A$ (the set of double singletons of elements of A) into one-to-one correspondence with the set of natural numbers $\{1, \ldots, n\}$ of the type appropriate for $A \in n$ to make sense. This can be proved by induction on the number of elements in A. If A has

one element a, clearly there is a bijection between $\{\{\{a\}\}\}$ and $\{1\}$ (all that needs to be checked is that these objects are of the same type: the number 1 being considered satisfies $A \in 1$). Suppose that for all $A \in n$, $\iota^{\mathbf{2}}\text{“}A \sim \{1, \ldots, n\}$. We want to show that for any $B \in n+1$, $\iota\text{“}B \sim \{1, \ldots, n+1\}$. $B = A \cup \{x\}$ for some $A \in n, x \notin A$. There is a bijection f from $\iota^{\mathbf{2}}\text{“}A$ to $\{1, \ldots, n\}$ by inductive hypothesis. $f \cup \langle \{\{x\}\}, n+1\rangle$ is easily seen to witness the desired equivalence in size.

Von Neumann's Counting Theorem: For any natural number n,

$$\{m \in \mathbb{N} \mid m < n\} \in T^{\mathbf{2}}(n).$$

Discussion: This is true for the same reasons. It is not really a theorem of von Neumann, but it relates to his representation of the natural numbers.

Notice that these counting theorems could be written in entirely unexciting forms if we adopted the Reasonable Convention above. It would then be the responsibility of the reader to spot the type difference and insert the appropriate T operation. This would have to be done in order to *prove* either of these statements.

A fully abstract definition of multiplication would say that $\kappa \cdot \lambda$ is the size of the union of κ disjoint sets each of size λ. To state this precisely requires the T operation just introduced.

***Definition (abstract definition of multiplication):** $\kappa \cdot \lambda$ is the uniquely determined cardinal of a set $\bigcup C$ where C is pairwise disjoint, $C \in T(\kappa)$, and $C \subseteq \lambda$.

The details of making this definition work are quite laborious. Infinity is required to show that there are such sets for any κ and λ, and Choice is required to show that the cardinal is uniquely determined. We regretfully eschew this definition and use a more concrete definition employing the cartesian product:

Definition (concrete definition of multiplication): $|A| \cdot |B|$ is defined as $|A \times B|$. It is straightforward to show that this does not depend on the choice of representatives A, B from the cardinals.

If we were using the Kuratowski pair we would define

$$|A| \cdot |B| = T^{-\mathbf{2}}(|A \times B|).$$

It would be harder to show that multiplication is total. We would also have

$$|A| + |B| = T^{-\mathbf{2}}(|(A \times \{0\}) \cup (B \times \{1\})|)$$

if we were using the Kuratowski pair.

The T operation commutes with arithmetic operations:

Theorem: For all cardinal numbers κ and λ, $T(\kappa) + T(\lambda) = T(\kappa + \lambda)$ and $T(\kappa \cdot \lambda) = T(\kappa) \cdot T(\lambda)$.

Theorems of cardinal arithmetic familiar from the theory of natural numbers (and from ordinary experience) have much more natural proofs in set theory than the inductive proofs given in Peano arithmetic.

Theorem: The following identities are true for all cardinal numbers κ, λ, μ (including natural numbers).

1. $\kappa + 0 = \kappa; \kappa \cdot 1 = \kappa$
2. $\kappa \cdot 0 = 0$
3. $\kappa + \lambda = \lambda + \kappa; \kappa \cdot \lambda = \lambda \cdot \kappa$
4. $(\kappa + \lambda) + \mu = \kappa + (\lambda + \mu); (\kappa \cdot \lambda) \cdot \mu = \kappa \cdot (\lambda \cdot \mu)$
5. $\kappa \cdot (\lambda + \mu) = \kappa \cdot \lambda + \kappa \cdot \mu$

All of these admit very natural proofs.

Sample Proofs:

commutativity of multiplication: Let κ, λ be cardinal numbers. Choose sets A and B such that $\kappa = |A|$ and $\lambda = |B|$. $\kappa \cdot \lambda = |A \times B|$ and $\lambda \cdot \kappa = |B \times A|$; what remains is to show that there is a bijection from $|A \times B|$ to $|B \times A|$. The map which sends each ordered pair $\langle a, b \rangle$ (for $a \in A, b \in B$) to $\langle b, a \rangle$ does the trick.

associativity of addition: Let κ, λ, μ be cardinal numbers. Choose A, B, C such that $\kappa = |A|, \lambda = |B|, \mu = |C|$. $(|A| + |B|) + |C| = |A \times \{0\} \cup B \times \{1\}| + |C| = |(A \times \{0\} \cup B \times \{1\}) \times \{0\} \cup C \times \{1\}| = |\{\langle\langle a, 0\rangle, 0\rangle \mid a \in A\} \cup \{\langle\langle b, 1\rangle, 0\rangle \mid b \in B\} \cup \{\langle c, 1\rangle \mid c \in C\}|$

Similarly $|A| + (|B| + |C|) = |\{\langle a, 0\rangle \mid a \in A\} \cup \{\langle\langle b, 0\rangle, 1\rangle \mid b \in B\} \cup \{\langle\langle c, 1\rangle, 1\rangle \mid c \in C\}|$.

A bijection from $\{\langle\langle a, 0\rangle, 0\rangle \mid a \in A\} \cup \{\langle\langle b, 1\rangle, 0\rangle \mid b \in B\} \cup \{\langle c, 1\rangle \mid c \in C\}$ to $\{\langle a, 0\rangle \mid a \in A\} \cup \{\langle\langle b, 0\rangle, 1\rangle \mid b \in B\} \cup \{\langle\langle c, 1\rangle, 1\rangle \mid c \in C\}$ is provided by the union of the map sending each $\langle\langle a, 0\rangle, 0\rangle$ to $\langle a, 0\rangle$, the map sending each $\langle\langle b, 1\rangle, 0\rangle$ to $\langle\langle b, 0\rangle, 1\rangle$ and the map sending each $\langle c, 1\rangle$ to $\langle\langle c, 1\rangle, 1\rangle$. Each of these maps is a bijection, they have disjoint domains and disjoint ranges, so their union is still a bijection. The existence of this bijection witnesses the desired equation.

Important arithmetic properties of the natural numbers *not* shared by general cardinals are the *cancellation properties*. It is not true in general that $\kappa + \mu = \lambda + \mu \leftrightarrow \kappa = \lambda$, nor that $\kappa \cdot \mu = \lambda \cdot \mu \wedge \mu \neq 0 \rightarrow \kappa = \lambda$. This means that we do not get sensible notions of subtraction or division.

But the following is a

Theorem: For any cardinals κ, λ and any natural number n, $\kappa + n = \lambda + n \rightarrow \kappa = \lambda$.

Proof: It suffices to prove $\kappa + 1 = \lambda + 1 \rightarrow \kappa = \lambda$: the result then follows by induction.

Suppose $\kappa + 1 = \lambda + 1$. Let A and B be chosen so that $\kappa = |A|$, $\lambda = |B|$, and neither A nor B is the universal set V. Note that if either A or B were the universal set, we could replace it with $V \times \{0\} \sim V$. Choose $x \notin A, y \notin B$. We have $|A \cup \{x\}| = \kappa + 1 = \lambda + 1 = |B \cup \{y\}|$. This means we can choose a bijection $f : (A \cup \{x\}) \rightarrow (B \cup \{y\})$. Either $f(x) = y$ or $f(x) \neq y$. If $f(x) = y$, then $f \lceil A$ is the desired bijection from A to B, witnessing $\kappa = \lambda$. If $f(x) \neq y$, then $f - \{\langle x, f(x)\rangle\} - \{\langle f^{-1}(y), y\rangle\} \cup \{\langle f^{-1}(y), f(x)\rangle\}$ is the desired bijection from A to B witnessing $\kappa = \lambda$. In either case we have established the desired conclusion.

2.11.1 Exercises

1. Prove that $|\mathbb{N}| + 1 = |\mathbb{N}| + |\mathbb{N}| = |\mathbb{N}| \cdot |\mathbb{N}| = |\mathbb{N}|$.

   Describe bijections by arithmetic formulas where you can; in any case clearly describe how to construct them (these are all familiar results, or should be, and all of the bijections can in fact be described algebraically: the formula for triangular numbers can be handy for this). I'm looking for bijections with domain $\mathbb{N}$ and range some more complicated set in every case.

2. Verify the distributive law of multiplication over addition in cardinal arithmetic,
   $$|A| \cdot (|B| + |C|) = |A| \cdot |B| + |A| \cdot |C|,$$
   by writing out explicit sets with the two cardinalities (fun with cartesian products and labelled disjoint unions!) and explicitly describing the bijection sending one set to the other. You do not need to prove that it is a bijection: just describe the sets and the bijection between them precisely.

3. Prove that $|\langle A, B\rangle| = |A| + |B|$ if the pair is taken to be a Quine pair.

4. Explain why the relation $A \sim B$ of equinumerousness (equipotence, being the same size) is an equivalence relation by citing basic properties of bijections.

   The structure of your proof should make it clear that you understand what an equivalence relation is.

   You do not need to prove the basic properties of bijections that are needed; you need only state them.

   Your proof should also make it clear that you know what $A \sim B$ means.

   What are the equivalence classes under the relation $\sim$ called in type theory?

5. In this problem you will indicate a proof of the associative property of multiplication for cardinal numbers.

   Recall that $|A| \cdot |B|$ is defined as $|A \times B|$.

   The goal is to prove that $(|A| \cdot |B|) \cdot |C| = |A| \cdot (|B| \cdot |C|)$. Describe sets of these cardinalities and (carefully) describe a bijection between them. You do not need to prove that the map is a bijection.

6. This may be so easy that it is hard.

   Prove the lemma stated in the text without proof that for functions f and g, $f \cup g$ is a function iff $f\lceil(\mathtt{dom}(f) \cap \mathtt{dom}(g)) = g\lceil(\mathtt{dom}(f) \cap \mathtt{dom}(g))$.

   This looks to me worth doing: it is an exercise in carefully unpacking definitions and keeping track of what it is that you need to show.

7. Prove that for m, n natural numbers, the definitions of $m+n$ and $m \cdot n$ given here coincide with the definitions based on iteration given earlier. These are induction proofs, of course.

2.12 Number Systems

In this section we give a development of the system of real numbers from the typed theory of sets. Part of the point is that this development is not unique or canonical in any way: we indicate how alternative developments might go. The development is full in the sense that all definitions of mathematical structures are given. Not all theorems are proved, though important ones are stated.

We begin with the system $\mathbb{N}^+$ of all nonzero natural numbers. We have already defined arithmetic operations of addition and multiplication on the natural numbers, and it is easy to see that $\mathbb{N}^+$ is closed under these operations.

We now give a construction of the system $\mathbb{Q}^+$ of *fractions* (positive rational numbers).

Definition: For $m, n \in \mathbb{N}^+$, we define $m|n$ as $(\exists x \in \mathbb{N}^+ . m \cdot x = n)$. This is read "$n$ is divisible by m" and we say that m is a *factor* of n.

Definition: For $m, n \in \mathbb{N}^+$, we define $\mathtt{gcd}(m, n)$ as the largest natural number x which is a factor of m and a factor of n. If $\mathtt{gcd}(m, n) = 1$, we say that m and n are *relatively prime*.

Theorem: If $m \cdot x = m \cdot y$, then $x = y$, where $m, x, y \in \mathbb{N}^+$.

Definition: If $m \cdot x = n$, we define $\frac{n}{m}$ as x (this is uniquely determined, if defined, by the previous theorem). Note that this notation will be superseded after the following definition.

Definition: We define a *fraction* as an ordered pair $\langle m, n \rangle$ of nonzero natural numbers such that m and n are relatively prime. For any ordered pair $\langle m, n \rangle$ of nonzero natural numbers, we define $\mathtt{simplify}(m, n)$ as $\left\langle \frac{m}{\mathtt{gcd}(m,n)}, \frac{n}{\mathtt{gcd}(m,n)} \right\rangle$. Note that $\mathtt{simplify}(m, n)$ is a fraction. After this point, we use the notation $\frac{m}{n}$ to denote $\mathtt{simplify}(m, n)$.

Observation: It is more usual to define an equivalence relation $\langle m, n \rangle \sim \langle p, q \rangle$ on ordered pairs of nonzero natural numbers (usually actually ordered pairs of integers with nonzero second projection) as holding when $mq = np$ (a proof that this is an equivalence relation is needed) then define fractions (more usually general rationals) as equivalence

classes under this relation. The construction given here uses canonical representatives instead of equivalence classes.

Definition: We define $\frac{m}{n} + \frac{p}{q}$ as $\frac{mq+np}{pq}$ and $\frac{m}{n} \cdot \frac{p}{q} = \frac{mp}{nq}$. We define $\frac{m}{n} \leq \frac{p}{q}$ as holding iff $mq \leq np$. We leave it to the reader to prove that these definitions are valid (do not depend on the choice of representation for the fractions), that $\leq$ is a linear order, and that addition and multiplication of fractions have expected properties. The complete familiarity of these definitions may obscure the fact that work needs to be done here.

Now we proceed to define the system of *magnitudes* (positive real numbers).

Definition: A *magnitude* is a set m of fractions with the following properties.

1. m and $\mathbb{Q}^+ - m$ are nonempty.
2. $(\forall pq \in \mathbb{Q}^+.p \in m \wedge q \leq p \rightarrow q \in m)$: m is downward closed.
3. $(\forall p \in m.(\exists q \in m.p \leq q))$: m has no largest element.

The motivation here is that for any positive real number r (as usually understood prior to set theory) the intersection of the interval $(0, r)$ with the set of positive rationals uniquely determines r (and of course is uniquely determined by r) and any set of positive rationals m with the properties given above will turn out to be the intersection of the set of positive rationals and $(0, \sup m)$.

Definition: For magnitudes m and n, we define $m + n$ as

$$\{p + q \mid p \in m \wedge q \in n\}$$

and $m \cdot n$ as

$$\{p \cdot q \mid p \in m \wedge q \in n\}.$$

We define $m \leq n$ as $m \subseteq n$. We leave it to the reader to prove that addition and multiplication of magnitudes always yield magnitudes and that these operations and the order relation have the expected properties.

This is where the payoff of our particular approach is found. It is more usual to use intersections of intervals $(-\infty, r)$ with all the rationals (positive, negative and zero) to represent the reals; with this representation of reals the definition of multiplication is horrible.

We cite a

Theorem: If $m + x = m + y$ then $x = y$, for magnitudes m, x, y.

Definition: If $m + x = n$, we define $n - m$ as x (uniqueness of $n - m$ if it exists follows from the previous theorem). This definition will be superseded by the following definition.

Definition: We define a *real number* as an ordered pair of magnitudes one of which is equal to 1 (where the magnitude 1 is the set of all fractions less than the fraction $1 = \frac{1}{1}$). For any pair of magnitudes $\langle x, y\rangle$, we define $\texttt{simp}(x, y)$ as $\langle (x + 1) - \min(x, y), (y + 1) - \min(x, y)\rangle$. Notice that $\texttt{simp}(x, y)$ will be a real number. Denote $\texttt{simp}(x, y)$ by $x - y$ (superseding the previous definition).

Definition: We define $(x - y) + (u - v)$ as $(x + u) - (y + v)$. We define $(x-y)\cdot(u-v)$ as $(xu+yv)-(xv+yu)$. We define $x-y \leq u-v$ as holding precisely when $x+v \leq y+u$. We leave it to the reader to establish that everything here is independent of the specific representation of $x - y$ and $u - v$ used, and that the operations and the order relation have expected properties.

A considerable amount of overloading is found here. Addition, multiplication and order are already defined for nonzero natural numbers when we start. In each system, addition, multiplication, and order are defined: these are different operations and relations in each system. Names of nonzero natural numbers, fractions, and magnitudes are also overloaded: the natural number n is confused with the fraction $\frac{n}{1}$ but it is not the same object, and similarly the magnitude $\{q \in \mathbb{Q}^+ \mid q < p\}$ is not the same object as the fraction p (and is one type higher than p!), and the real number $(m+1)-1$ is not the same object as the magnitude m, though in each case we systematically confuse them.

Certain important subsystems do not have a place in our development though they do in more usual developments.

Definition: We define the real number 0 as $1 - 1$. For each real number $r = x - y$ we define $-r$ as $y - x$. We define $r - s$ as $r + (-s)$ for reals r and s.

Definition: We define the set of *integers* $\mathbb{Z}$ as the union of the set of all (real numbers identified with) nonzero naturals, $\{0\}$, and the set of all additive inverses $-n$ of (real numbers identified with) nonzero naturals n.

Definition: We define the set of *rationals* $\mathbb{Q}$ as the union of the set of all (real numbers identified with) fractions p, $\{0\}$, and the set of all additive inverses $-p$ of (real numbers identified with) fractions p.

Definition: For any fraction $q = \frac{m}{n}$ we define q^{-1} as $\frac{n}{m}$. For any magnitude m, we define m^{-1} as $\{q^{-1} \mid q \notin m\}$. It is straightforward to prove that m^{-1} is a magnitude and $m \cdot m^{-1} = 1$ for each m. Now define the reciprocal operation for reals: $((m + 1) - 1)^{-1} = (m^{-1} + 1) - 1$ and $(1 - (m + 1))^{-1} = 1 - (m^{-1} + 1)$ for each magnitude m, while $(1 - 1)^{-1}$ is undefined. It can be proved that $r \cdot r^{-1} = 1$ for each real $r \neq 0$. Finally, we define $\frac{r}{s}$ as $r \cdot s^{-1}$ for any real r and nonzero real s.

We noted above that we have avoided the use of equivalence classes of ordered pairs at the steps passing to fractions and to signed real numbers, preferring to use canonical representatives. Simplification of fractions is of course a familiar mathematical idea; the canonical representation of reals we use is less obvious but works just as well.

In this development we have followed the prejudices of the ancient Greeks as far as possible, delaying the introduction of zero or negative quantities to the last step.

The reals as defined here satisfy the following familiar axioms of a "complete ordered field". Up to a suitable notion of isomorphism, the reals are the only complete ordered field.

commutative laws: $a + b = b + a$; $a \cdot b = b \cdot a$.

associative laws: $(a + b) + c = a + (b + c)$; $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

distributive law: $a \cdot (b + c) = a \cdot b + a \cdot c$.

identity laws: $a + 0 = a$; $a \cdot 1 = a$.

inverse laws: $a + (-a) = 0$; $a \cdot a^{-1} = 1$ if $a \neq 0$.

nontriviality: $0 \neq 1$

closure of positive numbers: If $a \geq 0$ and $b \geq 0$ then $a + b \geq 0$ and $a \cdot b \geq 0$. [note that $a \geq 0$ is a primitive notion at this point in the development: the reals of the form $r = (m + 1) - 1$ are the ones for which we assert $r \geq 0$].

trichotomy: For each real number a, exactly one of the following is true: $a \geq 0$, $a = 0$, $-a \geq 0$.

Definition: $a \leq b$ iff $b + (-a) \geq 0$.

Theorem: $\leq$ thus defined is a linear order.

completeness: Any nonempty set of reals which is bounded above (in terms of the order just defined) has a least upper bound.

2.12.1 Exercises

1. Show as many of the properties of the real number system stated as the end of the section true (prove them) or false (exhibit a counterexample) as you can for

   (a) the system of fractions

   (b) the system of magnitudes

   (c) the integers

   (d) the rational numbers

   (e) the entire system of reals

   This is an altogether unreasonable question!

2.13 Well-Orderings and Ordinal Numbers

We recall that a well-ordering is a linear order with the property that the corresponding strict partial order is well-founded.

Definition: A *well-ordering* is a linear order $\leq$ with the property that for each nonempty subset A of $\texttt{fld}(\leq)$ there is $a \in A$ such that there is no $b \neq a$ in A such that $b \leq a$: such an a is a minimal element of A (in fact, the minimal element is unique because $\leq$ is linear).

In this section, we study the structure of well-orderings. In this section we state and prove powerful and highly abstract theorems: for some concrete discussion of ordinal numbers, look toward the end of the next section.

Definition: Two relations R and S are said to be *isomorphic* iff there is a bijection f from $\texttt{fld}(R)$ to $\texttt{fld}(S)$ such that for all x, y, $x\,R\,y \leftrightarrow f(x)\,S\,f(y)$. f is said to be an *isomorphism* from R to S. We write $R \approx S$ for "R is isomorphic to S".

Theorem: Isomorphism is an equivalence relation on relations.

Definition: An equivalence class under isomorphism is called an *isomorphism type*.

Definition: Well-orderings are said to be *similar* iff they are isomorphic.

Theorem: A relation isomorphic to a well-ordering is a well-ordering.

Definition: The isomorphism type of a well-ordering is called its *order type*. We write $\texttt{ot}(\leq)$ for the order type $[\leq]_\approx$ of $\leq$. A set is an *ordinal number* iff it is the order type of some well-ordering. The set of all ordinal numbers is called `Ord`.

There are few well-orderings familiar to us from undergraduate mathematics. Any finite linear order is a well-ordering.

Theorem: For any $n \in \mathbb{N}$, any two linear orders with field of size n are isomorphic and are well-orderings.

Theorem: A well-ordering is finite iff its converse is also a well-ordering.

Our use of "finite" in the previous theorem might cause confusion, which will be alleviated by considering the following

Lemma: A relation (considered as a set) is finite iff its field is finite.

Definition (finite ordinals): For each natural number n, there is a unique ordinal number which is the order type of all orders with range of that cardinality: we also write this ordinal number as n, though it is not the same object as the cardinal number n.

An amusing observation, depending crucially on the exact details of our implementation, is the following relationship between ordinal and cardinal numbers.

Theorem: The ordinal number n is a subset of the cardinal number $\frac{n(n-1)}{2}$.

In the usual untyped set theory, with the usual implementations of the notions of ordinal and cardinal number, the finite cardinals and the finite ordinals are the same objects. We will see this in chapter 4.

The usual order on the natural numbers is a well-ordering. The usual orders on the integers, rationals and reals are *not* well-orderings. Another example of an infinite well-ordering which is familiar from calculus is the order on reals restricted to the range of a strictly increasing bounded convergent sequence taken together with its limit.

Definition: We define ω as the order type of the natural order on the natural numbers.

We give some basic definitions for arithmetic of ordinal numbers.

Definition (ordinal addition): For well-orderings R and S, we define another well-ordering $R \oplus S$. The field of $R \oplus S$ is $\mathtt{fld}(R) \times \{0\} \cup \mathtt{fld}(S) \times \{1\}$. $\langle x, i\rangle\,(R \oplus S)\,\langle y, j\rangle$ is defined as $i < j \vee i = 0 \wedge j = 0 \wedge x\,R\,y \vee i = 1 \wedge j = 1 \wedge x\,S\,y$. Intuitively, we make disjoint orders of types R and S and put the order of type R in front of the order of type S. Finally, we define $\alpha + \beta$ for ordinals α and β as $\mathtt{ot}(R \oplus S)$ for any $R \in \alpha$ and $S \in \beta$.

Another way to put this: for any relation R, define R_x as $\{\langle\langle a, x\rangle, \langle b, x\rangle\rangle \mid a\,R\,b\}$. Notice that $R \approx R_x$ for any R and x and $R_x \cap S_y = \emptyset$ for any R and S and any distinct x and y. For any ordinals α, β define $\alpha + \beta$ as $\mathtt{ot}(R_0 \cup (\mathtt{fld}(R_0) \times \mathtt{fld}(S_1)) \cup S_1)$ where $R \in \alpha$ and $S \in \beta$. It is straightforward to establish that $R_0 \cup (\mathtt{fld}(R_0) \times \mathtt{fld}(S_1)) \cup S_1$ is a well-ordering and that its order type does not depend on which representatives R and S are chosen from α and β.

Discussion: An order of type $\omega + 1$ is readily obtained: define $x \leq' y$ as

$$x \in \mathbb{N} \wedge y \in \mathbb{N} \wedge 0 < x \leq y \vee y = 0.$$

In effect, we move 0 from its position at the beginning of the order to the end. This is the same order type as that of a strictly increasing sequence taken together with its limit, which we mentioned above.

The relation $\leq'$ is not isomorphic to the usual $\leq$ on the natural numbers. An easy way to see this is that there is a $\leq'$-largest element of the field of $\leq'$, and this is a property of relations which is preserved by isomorphism: if $\leq' \approx \leq$ were witnessed by an isomorphism f then $f(0)$ would have to be the $\leq$-largest natural number, and there is no such natural number.

Further, the field of $\leq'$ is the same size as the field of $\leq$ (in fact, it is the same set!): so the theorem that there is a unique order type of well-orderings of each finite cardinality n does not generalize to infinite cardinalities.

Observe that an order of type $\omega+\omega$ is a still more complex well-ordering with field the same size as the field of a relation of type ω. A concrete example of such an order would be the order

$$\{\langle x, y\rangle \in \mathbb{N}^2 \mid 2|(x - y) \wedge x \leq y \vee 2 \not| x \wedge 2|y\},$$

which puts the odd and even numbers in their usual respective orders but puts all the odd numbers before all the even numbers.

Definition (ordinal multiplication): For well-orderings R and S, we define another well-ordering $R \otimes S$. The field of $R \otimes S$ is $\mathtt{fld}(R) \times \mathtt{fld}(S)$. $\langle x, y\rangle \, (R \otimes S) \, \langle u, v\rangle$ is defined as $y \, S \, v \vee y = v \wedge x \, R \, u$. This is reverse lexicographic order on the cartesian product of the fields of the relations. Finally, we define $\alpha \cdot \beta$ for ordinals α and β as $\mathtt{ot}(R \otimes S)$ for any $R \in \alpha$ and $S \in \beta$.

The order $\omega \cdot \omega$ is a still more complex order type whose field is the same size as that of any relation of order type ω. There are very complicated well-orderings with countable fields (whose order types are called *countable ordinals*).

The algebra of ordinal numbers contains surprises. Some algebraic laws do work much as expected, but some basic laws are not inherited from the

algebra of natural numbers. For example, $\omega + 1 \neq 1 + \omega = \omega$ and $\omega \cdot 2 \neq 2 \cdot \omega = \omega$.

We now study the natural order relation on the ordinal numbers, which turns out to be a well-ordering itself (at a higher type).

Definition: If $\leq$ is a partial order and $x \in \mathtt{fld}(\leq)$, we define $\mathtt{seg}_{\leq}(x)$ as $\{y \mid y < x\}$ (where $<$ is the strict partial order $[\leq] - [=]$). $\mathtt{seg}_{\leq}(x)$ is the *segment* determined by x. We define $\leq_x$ as $[\leq] \cap \mathtt{seg}_{\leq}(x)^2$; this is the *segment restriction* of $\leq$ determined by x.

Theorem: If $\leq$ is a well-ordering and $x \in \mathtt{fld}(\leq)$ then $\leq_x$ is a well-ordering.

Lemma: No well-ordering is isomorphic to one of its own segment restrictions.

Proof: Suppose that $\leq$ is a well-ordering, x is in the field of $\leq$, and $\leq \approx (\leq)_x$ is witnessed by an isomorphism f. Since $f(x) \neq x$ is obvious (x is not in the range of f!), there must be a $\leq$-least y such that $f(y) \neq y$. Let $A = \mathtt{seg}_{\leq}(y)$. Each element of A is fixed by f. In $\leq$, y is the least object greater than all elements of A. In $(\leq)_x$, $f(y)$ is the least object greater than all elements of A. The two orders agree on the common part of their field. Since $f(y)$ is certainly in the field of $\leq$, we have $y \leq f(y)$ (as otherwise $f(y)$ would be a smaller strict upper bound for A in $\leq$). Since $y \leq f(y)$, we have y in the field of $(\leq)_x$, and $f(y) \leq y$, as otherwise y would be a smaller strict upper bound for A in $(\leq)_x$. So $y = f(y)$, which is a contradiction.

Corollary: No two distinct segment restrictions of the same well-ordering can be isomorphic to one another.

Proof: One of them would be a segment restriction of the other.

Definition: We say that a subset D of the field of a well-ordering $\leq$ is "downward closed in $\leq$" iff $(\forall d \in D.(\forall e \leq d.e \in D))$.

Lemma: For any well-ordering $\leq$, a set downward closed in $\leq$ is either the field of $\leq$ or a segment in $\leq$.

Proof: Let D be a set downward closed in $\leq$. If x belongs to the field of $\leq$ but does not belong to D, then $d < x$ must be true for all $d \in D$,

as otherwise we would have $x \leq d \in D$, from which $x \in D$ would follow. This means that if D has no strict upper bound, it must be the entire field of $\leq$. If D does have a strict upper bound, it must have a $\leq$-least strict upper bound x because $\leq$ is a well-ordering. We claim that $D = \mathtt{seg}_{\leq}(x)$ in this case. If $y \in \mathtt{seg}_{\leq}(x)$, then y cannot be a strict upper bound of D because x is the least strict upper bound of D, and so y must be an element of D. If y is an element of D, then y must be less than x because x is a strict upper bound of D, that is, y is an element of $\mathtt{seg}_{\leq}(x)$. Sets with the same elements are the same.

Theorem: If $\leq_1$ and $\leq_2$ are well-orderings, then exactly one of three things is true: either $\leq_1$ and $\leq_2$ are isomorphic, or $\leq_1$ is isomorphic to a segment restriction $(\leq_2)_x$, or $\leq_2$ is isomorphic to a segment restriction $(\leq_1)_x$.

Proof: Let $\leq_1$ be a well-ordering with field A. Let $\leq_2$ be a well-ordering with field B. Define C as $\{a \in A \mid \neg(\exists b \in B.(\leq_1)_a \approx (\leq_2)_b)\}$, the set of all elements of the field of $\leq_1$ whose segment restrictions are *not* isomorphic to a segment restriction in $\leq_2$. If C is nonempty, it has a least element c. Each $d <_1 c$ does not belong to C, because c is the $\leq_1$-least element of C. Thus, by the definition of C, there is an $e \in B$ such that $(\leq_1)_d \approx (\leq_2)_e$. There can be only one such e because no two segment restrictions of the same well-ordering can be isomorphic to each other. Thus there is a function F which maps each $d <_1 c$ to the unique e such that $(\leq_1)_d \approx (\leq_2)_e$. We claim that F is an isomorphism from $(\leq_1)_c$ to $\leq_2$. This breaks down into three subclaims: F is an injection, F is order-preserving, and the range of F is B. For each $d <_1 c$, we have an isomorphism f witnessing $(\leq_1)_d \approx (\leq_2)_{F(d)}$. For each $d' < d$, the restriction of f to $\mathtt{seg}_{\leq_1}(d')$ is an isomorphism from $(\leq_1)_{d'}$ to $(\leq_2)_{f(d')}$, so in fact $F(d') = f(d')$. Because the range of f is the segment in $\leq_2$ determined by $F(d)$, we have $F(d') = f(d') < F(d)$. This shows both that F is order preserving and that it is a bijection. Further, it shows that the range of F is downward closed, as we see that the restriction of F to the segment determined by d is the isomorphism from the segment determined by d to the segment determined by $F(d)$. Since the range of F is downward closed, it must be either B or some $\mathtt{seg}_{\leq_2}(x)$, so F is either an isomorphism from $(\leq_1)_c$ to $\leq_2$ or an isomorphism from $(\leq_1)_c$ to some $(\leq_2)_x$. The latter case is impossible by the definition

of c, so we must actually have F an isomorphism from $(\leq)_c$ to $\leq_2$, establishing the Theorem in this case. If the set C is empty, then for every $a \in A$ there is $b \in B$ such that $(\leq_1)_a \approx (\leq_2)_b$. This b must be unique as no two distinct segment restrictions of $\leq_2$ can be isomorphic. For each $a \in A$, we define $F(a)$ as the unique b such that $(\leq_1)_a \approx (\leq_2)_b$. Exactly the same argument just given shows that F is a bijection, order-preserving, and has a downward closed range. From this it follows just as in the first case that F is an isomorphism from $\leq_1$ to either $\leq_2$ or some $(\leq_2)_x$, establishing that the Theorem is true in this case. If $\leq_1 \approx \leq_2$ then we cannot have either $(\leq_1)_x \approx \leq_2$ or $(\leq_2)_x \approx \leq_1$ because a well-ordering cannot be similar to one of its segment restrictions. If we had $\leq_1 \approx (\leq_2)_x$, and further had $\leq_2 \approx (\leq_1)_y$, witnessed by an isomorphism g, then we would have $\leq_1 \approx (\leq_1)_{g(x)}$, which is impossible. This establishes that only one of the three cases can hold.

Definition: If α and β are ordinal numbers, we define $\alpha \leq \beta$ as holding iff either $\alpha = \beta$ or each element of α is isomorphic to a segment restriction in each element of β.

Theorem: The relation $\leq$ defined on ordinal numbers in the previous definition is a well-ordering. Where it is necessary to distinguish it from other orders, we write it $\leq_\Omega$.

Proof: Let α and β be ordinals. If $\leq_1 \in \alpha$ and $\leq_2 \in \beta$, then either $\leq_1 \approx \leq_2$, in which case $\alpha = \beta$, or $\leq_1$ is isomorphic to a segment restriction in $\leq_2$, in which case the same is true for any $\leq_1' \approx \leq_1$ and $\leq_2' \approx \leq_2$, or $\leq_2$ is isomorphic to a segment restriction in $\leq_1$, in which case the same is true for any $\leq_2' \approx \leq_2$ and $\leq_1' \approx \leq_1$. If more than one of these alternatives held for any pair of well-orderings, one of them could be shown to be isomorphic to one of its own segment restrictions. Certainly $\alpha \leq \alpha$, so the $\leq$ relation on ordinals is reflexive. If $\alpha \leq \beta$ and $\beta \leq \alpha$ this must be witnessed by isomorphisms between $\leq_1 \in \alpha$ and $\leq_2 \in \beta$ in both directions, or once again we would have one of these well-orderings isomorphic to a segment restriction of itself. So the $\leq$ relation on ordinals is anti-symmetric. If we have $\alpha \leq \beta$ and $\beta \leq \gamma$ and we choose $\leq_1, \leq_2, \leq_3$ in α, β, γ respectively, we have $\leq_1$ isomorphic to $\leq_2$ or a segment restriction thereof, and $\leq_2$ isomorphic to $\leq_3$ or a segment restriction thereof, and composition of isomorphisms gives us an isomorphism from $\leq_1$ to $\leq_3$ or a segment restriction thereof, thus $\alpha \leq \gamma$, so the $\leq$ relation on

ordinals is transitive and is a linear order. Now let $\mathcal{A}$ be a nonempty set of ordinals. Let $\alpha \in \mathcal{A}$. Let $\leq_1 \in \alpha$ have field A. Consider the set B of all $a \in A$ such that $(\leq_1)_a$ belongs to some element of $\mathcal{A}$. If B is empty, then α is the $\leq$-smallest element of $\mathcal{A}$. If B is nonempty, choose the smallest a in B: $\mathtt{ot}((\leq_1)_a)$ is the $\leq$-smallest element of $\mathcal{A}$. So the relation $\leq$ on the ordinal numbers is a well-ordering, which is what we set out to prove.

Definition: $\mathtt{ot}(\leq_\Omega)$ is called Ω: notice that Ω is not of the same type as the ordinals in the field of the relation $\leq_\Omega$ of which it is the order type (it is 2 types higher; it would be 4 types higher if we defined well-orderings using the Kuratowski pair).

2.13.1 Exercises

1. Some linear orders are listed. For each one, state (correctly) that it is a well-ordering or that it is not. If it is not, explain precisely why it is not (this means give an example of something). If it is, give its order type (an ordinal number).

   (a) $\emptyset$

   (b) the standard order on the integers restricted to $\{x \in \mathbb{Z} \mid -2 \leq x \leq 2\}$

   (c) the standard order on the integers restricted to $\{x \in \mathbb{Z} \mid x \leq 0\}$

   (d) the standard order on the rationals restricted to $\{\frac{n}{n+1} \mid n \in \mathbb{N}\} \cup \{1\}$

   (e) the standard order on the rationals restricted to $\{\frac{n+1}{n} \mid n \in \mathbb{N}\} \cup \{1\}$

   (f) the standard order on the reals restricted to the interval $[0, 1]$

2. Prove that for any natural number n, any two linear orders with a field of size n are isomorphic, and all such linear orders are well-orderings. (How do we prove anything about natural numbers?)

3. Prove that if R and S are well-orderings, so is $R \oplus S$. You need to prove that it is a linear order (which will probably require some reasoning by cases) and prove that it has the additional defining property of a well-ordering.

   Now that you are filled with self-confidence, do the same for $R \otimes S$.

4. Define sets of real numbers such that the restriction of the standard order on the real numbers to that set has each of the following order types:

   (a) $\omega + 1$

   (b) $\omega \cdot 3$

   (c) $3 \cdot \omega$

   (d) $\omega \cdot \omega$

   (e) $\omega \cdot \omega \cdot \omega$ (OK I suppose this is nasty, but see if you can do it)

5. Prove your choice of the two following annoying propositions (these are annoying in the sense that they are straightforward (even "obvious") but there is a good deal to write down).

   (a) Isomorphism is an equivalence relation on relations.

   (b) A relation isomorphic to a well-ordering is a well-ordering.

2.14 Transfinite Induction and Recursion

The following theorem is an analogue of mathematical induction for the ordinals.

Transfinite Induction Theorem: Suppose A is a set of ordinals with the following property: $(\forall\alpha \in \mathtt{Ord}.(\forall\beta < \alpha.\beta \in A) \rightarrow \alpha \in A)$. Then $A = \mathtt{Ord}$.

Proof: If $A \neq \mathtt{Ord}$, then $\mathtt{Ord} - A$ is a nonempty set and so contains a least ordinal α. But then obviously $(\forall\beta < \alpha.\beta \in A)$, so $\alpha \in A$ by assumption, which is a contradiction.

Transfinite Induction Theorem (bounded form): Suppose A is a set of ordinals with the following property: $(\forall\alpha < \gamma.(\forall\beta < \alpha.\beta \in A) \rightarrow \alpha \in A)$. Then $(\forall\alpha < \gamma.\alpha \in A)$.

Transfinite Induction Theorem (property form): Suppose $\phi[\alpha]$ is a formula such that $(\forall\alpha \in \mathtt{Ord}.(\forall\beta < \alpha.\phi[\beta]) \rightarrow \phi[\alpha])$. Then $(\forall\alpha \in \mathtt{Ord}.\phi[\alpha])$.

This looks like the theorem of strong induction for the natural numbers. We can make it look a bit more like the usual formulation of induction by defining some operations on ordinals. The alternative forms are easy to prove and are relevant to untyped set theory where there is no set containing all ordinals. [The property form would have to be restated using a predicate $\mathtt{Ord}(x)$ in place of a set of all ordinals to prove theorems about all ordinals in a context where there is no *set* of all ordinals.]

zero: We define 0 as the smallest ordinal (the order type of the empty well-ordering).

successor: For any ordinal α, we define the *successor* of α as the smallest ordinal greater than α. No special notation is needed for successor, since it is easy to show that the successor of α is $\alpha + 1$. Every ordinal has a successor: for any infinite ordinal α containing a well-ordering W with minimal element x, $\mathtt{ot}(W - (\{x\} \times \mathtt{fld}(W)) \cup (\mathtt{fld}(W) \times \{x\}))$ is $\alpha + 1$: the new order is obtained by moving the minimal element of W from bottom to top of the order.

limit ordinal: A nonzero ordinal which is not a successor is called a limit ordinal.

Now we give a different formulation of Transfinite Induction.

Transfinite Induction Theorem: Suppose that A is a set of ordinals such that $0 \in A$, for every ordinal $\alpha \in A$ we also have $\alpha + 1 \in A$, and for any limit ordinal λ such that for all $\beta < \lambda$ we have $\beta \in A$, we also have $\lambda \in A$. Then $A = \texttt{Ord}$.

Proof: Again, consider the smallest element of the complement of A (there must be a smallest if there is any). It cannot be 0 because $0 \in A$. It cannot be a successor (because its predecessor would be in A, so it would be in A). It cannot be a limit (because everything below it would be in A, so it would be in A). These are the only possibilities.

We now give an extended example of proof by transfinite induction. For purposes of this example, we assume familiarity with the real numbers at the usual undergraduate level. We have seen in an earlier section of these notes how to construct the real numbers in our type theory; mod omitted proofs we are warranted in assuming that they are available at some type and have familiar properties.

Definition: We say that an ordinal α is a *countable ordinal* iff the relations which belong to it have countably infinite fields.

Lemma: For any countable ordinal α, there is a function $f : \mathbb{N} \to \texttt{Ord}$ such that for natural numbers $i < j$ we have $f(i) < f(j)$, $f(i) < \alpha$ for all i, and α is the least ordinal greater than all $f(i)$'s. More briefly, f is a strictly increasing sequence of ordinals whose least upper bound is α. We will reserve the right to use the usual notation for sequences, writing $f(i) = \alpha_i$.

Proof of Lemma: Let α be a countable ordinal, and let $\leq$ be a fixed well-ordering of type α with field A. Because α is a countable ordinal, there is an enumeration a_i of the set A (the function $i \in \mathbb{N} \mapsto a_i$ being a bijection from $\mathbb{N}$ to A). We define a sequence b_i recursively as follows: $b_0 = a_0$. Once b_i has been defined as a_j, we define b_{i+1} as a_k, where k is the least natural number such that $a_j < a_k$. The sequence b is strictly

increasing, and every element of A is $\leq$-dominated by some element of the range of this sequence (because $a_k \leq b_k$ for every natural number k, as is easy to prove by induction). We can thus define $f(i) = \alpha_i$ as $\mathtt{ot}(\leq)_{b_i}$: these ordinals are clearly all less than the order type α of $\leq$, they increase strictly as the index increases, and any ordinal less than α, being the order type of some $(\leq)_{a_k}$, is dominated by some α_k.

Definition: For any subset X of the interval $(0, 1]$ in the reals and any $a < b$ real numbers, we define $X_{[a,b]}$ as $\{(1 - x)a + xb \mid x \in X\}$. This is a scaled copy of X in the interval $[a, b]$.

For any function f from $\mathbb{N}$ to $\mathcal{P}((0, 1])$ (infinite sequence of subsets of $(0, 1]$), define f^* as $\bigcup\{f(n)_{[1-2^{-n},1-2^{-n-1}]} \mid n \in \mathbb{N}\}$. This construction allows me to put together scaled copies of the infinite sequence of sets $f(n)$, so that the scaled copies are disjoint and appear in the same order that the sets appear in the sequence.

Theorem: For any finite or countable ordinal α, we can find a set of reals $A_\alpha \subseteq (0, 1]$ such that the order type of the restriction of the usual linear order on the reals to A_α is a well-ordering of order type α.

Proof: We break the proof into three cases: $\alpha = 0$, $\alpha = \beta + 1$ for some β, or α a limit ordinal.

In any of these cases, we assume that sets $A_\beta \subseteq (0, 1]$ of reals such that the usual order on the reals restricted to A_β has order type β exist for each ordinal $\beta < \alpha$. Our goal is to show we can find a set of reals A_α such that the order type of the restriction of the usual linear order on the reals to A_α is a well-ordering of order type α.

If $\alpha = 0$, $A_\alpha = \emptyset$ is a subset of the reals such that the restriction of the natural order on the reals to this set has order type $\alpha = 0$.

If $\alpha = \beta+1$, we assume the existence of A_β as above. The set $(A_\beta)_{[0,\frac{1}{2}]} \cup \{1\}$ has the desired properties: the order type of the natural order on the reals restricted to this set is clearly $\beta + 1$.

If α is a limit ordinal, we have two cases to consider. If α is not a countable ordinal, we have nothing to prove. If α is a countable ordinal, we select a strictly increasing sequence α_i such that the least upper bound of its range is α, as a Lemma above shows we are entitled to do. For each α_i, we are given a set $A_{\alpha_i} \subseteq (0, 1]$ of reals with associated

order type α_i. For each i, we select a subset A'_{α_i} of A_{α_i} which we now define. A'_{α_0} is defined as A_{α_0}. For each i, $\leq_{\mathbb{R}} \lceil A_{\alpha_{i+1}}$ has a unique segment restriction of order type α_i. $A'_{\alpha_{i+1}}$ is obtained by subtracting the field of this segment restriction from $A_{\alpha_{i+1}}$. Define $f(i)$ as A'_{α_i} and the set f^* will be the desired set: this is the union of linearly scaled copies of all the A'_{α_i}'s, made successively smaller so they will all fit into $(0, 1]$. It should be clear that the union of such linearly scaled copies has order type α.

Theorem: Any ordinal α which is the order type of the natural order on a subset A of the reals is finite or countable.

Proof: Given such an ordinal α and set A, we construct a set A' such that the natural order on A' also has order type α and all elements of A' are rational numbers (so A' must be finite or countable). For each element $a \in A$, either a is the largest element of A or there is a first element a' of A which is greater than a. This is true because A is well-ordered by the usual order on the reals. Assume that we have an enumeration q_i of the rationals. Let q_a be the first rational in this enumeration which is greater than a and less than a' (or simply the first rational in this enumeration which is greater than a, if a is the largest element of A). It should be evident for all $a, b \in A$ that $q_a < q_b \leftrightarrow a < b$. Thus $\{q_a \mid a \in A\}$ is a set of rationals (thus finite or countable) and the order type of the natural order on this set is α, so α is a finite or countable ordinal.

We conclude that the order types of well-orderings that we can construct as suborders of the natural order on the real numbers are exactly the finite and countable ordinals. We will see below that there are uncountable ordinals (this will be our first evidence that there are infinite sets which are not countably infinite).

We introduce a type raising operation on ordinals analogous to that already given for cardinals and also traditionally denoted by T.

Definition: For any relation R, we define R^ι as $\{\langle \iota(x), \iota(y)\rangle \mid x\,R\,y\} = \{\langle \{x\}, \{y\}\rangle \mid x\,R\,y\}$. Notice that R^ι is one type higher than R and would seem in some external sense to be isomorphic to R. $R^{\iota^{\mathbf{n}}}$ is similarly defined as $\{\langle \iota^{\mathbf{n}}(x), \iota^{\mathbf{n}}(y)\rangle \mid x\,R\,y\}$

Definition: For any ordinal α, we define $T(\alpha)$ as $\mathtt{ot}(R^\iota)$ for any $R \in \alpha$ (it is easy to show that the choice of R does not matter). Of course we can then also define $T^{\mathbf{n}}(\alpha)$ and $T^{-\mathbf{n}}(\alpha)$ in the natural ways.

Induction can actually be carried out along any well-ordering, but it is traditional to translate all transfinite inductions into terms of ordinals. A general way to do this involves indexing the elements of $\mathtt{fld}(\leq)$ for a general well-ordering $\leq$ with ordinals:

Definition(ordinal indexing): For any well-ordering W, define W_α as the unique element x of $\mathtt{fld}(W)$ (if there is one) such that $\mathtt{ot}((\leq)_x) = \alpha$. [Note that if W is a well-ordering of a set of ordinals this is different from $(W)_\alpha$, the segment restriction of W to elements which are W-less than α.]

Notice that the type of α is one higher than the type of W and two higher than the type of W_α (it would be four higher than the type of W_α if we used the Kuratowski pair).

W_α will be defined for each α iff $\alpha < \mathtt{ot}(W)$.

Discussion of ordinal indexing in the natural order on the ordinals themselves requires the following

Theorem: $\mathtt{ot}((\leq_\Omega)_\alpha) = T^{\mathbf{2}}(\alpha)$

Proof: This is proved by transfinite induction. Note that what it says is that the order type of the segment restriction of the natural order on the ordinals to the ordinals less than α is $T^{\mathbf{2}}(\alpha)$. It is "obvious" that this order type is actually α itself, but of course the order type of the segment restriction is two types higher than α itself, so it is seen to be the corresponding ordinal $T^{\mathbf{2}}(\alpha)$ two types higher.

So $[\leq_\Omega]_\alpha = T^{-\mathbf{2}}(\alpha)$ (not α itself).

Note that $[\leq_\Omega]_\alpha$ will be undefined for $\alpha = \mathtt{ot}(\leq_\Omega) = \Omega$, but $[\leq_\Omega]_{T^{\mathbf{2}}(\Omega)} = \Omega$. This shows that $T^{\mathbf{2}}(\Omega)$ is not equal to Ω: in fact $T^{\mathbf{2}}(\Omega) < \Omega$ because $T^{\mathbf{2}}(\Omega)$ is the order type of a segment restriction of the natural order on the ordinals, whose order type is Ω.

The result that $T^{\mathbf{2}}(\Omega) < \Omega$ (in which there is of course a kind of punning reference to the sets of ordinals at different types) shows that there are in effect more ordinals in higher types. There is no well-ordering in type k as long as the natural order on the ordinals in type $k+2$.

Now we prove that there are uncountable ordinals.

Theorem: There are ordinals which are not finite or countably infinite (in sufficiently high types), and so there is in particular a first uncountably infinite ordinal ω_1.

Proof: Consider the restriction of the natural well-ordering on the ordinals to the finite and countable ordinals. This is a well-ordering, so it has an order type, which we call ω_1. For each countable ordinal α, the order type of $(\leq_\Omega)_\alpha$ is $T^{\mathbf{2}}(\alpha)$, and of course $T^{\mathbf{2}}(\alpha) < \omega_1$, because the former is the order type of a segment restriction of the latter. So it cannot be the case that $\omega_1 = T^{\mathbf{2}}(\alpha)$ for any countable ordinal α (of type two lower than that of ω_1). It only remains to show that every countable ordinal of the same type as ω_1 is of the form $T^{\mathbf{2}}(\beta)$. Suppose that γ is a countable ordinal of the same type as ω_1. γ is the order type of some well-ordering $\leq$ with field the set of natural numbers. Now consider $\{\langle T^{-\mathbf{2}}(m), T^{-\mathbf{2}}(n)\rangle \mid m \leq n\}$. We know that there is a set of natural numbers two types lower than the one that $\leq$ orders, because γ is of the same type as ordinals $T^{\mathbf{2}}(\alpha)$ with α countable. We know that the $T^{-\mathbf{1}}$ operation is total on the natural numbers. It follows that the relation just defined makes sense and is of some countable order type β, with $\gamma = T^{\mathbf{2}}(\beta)$, so $\gamma < \omega_1$. But γ is an arbitrary countable ordinal of the type of ω_1, so ω_1 is uncountably infinite.

Corollary: There are sets which are infinite but not countably infinite.

Proof: The field of any relation of type ω_1 will serve: the set of finite and countable ordinals is shown to be uncountably infinite in the argument above.

Here is another very important result about well-orderings whose proof is assisted by ordinal indexing.

Theorem: Suppose that $\leq_1 \subseteq \leq_2$ are well-orderings. Then $\mathtt{ot}(\leq_1) \leq \mathtt{ot}(\leq_2)$.

Proof: We can prove by an easy transfinite induction that $[\leq_2]_\alpha$ is defined and $[\leq_2]_\alpha \leq_2 [\leq_1]_\alpha$ for each ordinal $\alpha < \mathtt{ot}(\leq_1)$. The map taking each $[\leq_1]_\alpha$ to $[\leq_2]_\alpha$ is the desired isomorphism witnessing $\mathtt{ot}(\leq_1) \leq \mathtt{ot}(\leq_2)$.

Of course, when the author says something is easy, that means he or she doesn't really want to take the trouble to prove it. We now do so.

We prove by transfinite induction that $[\leq_2]_\alpha$ is defined and $[\leq_2]_\alpha \leq_2 [\leq_1]_\alpha$ for each ordinal $\alpha < \mathtt{ot}(\leq_1)$.

Note first that an ordinal α is less than $\mathtt{ot}(\leq_1)$ precisely if it is the order type of some $(\leq_1)_x$, by the definition of the order on the ordinals, and this x is $[\leq_1]_\alpha$ by the definition of ordinal indexing, so certainly $[\leq_1]_\alpha$ is defined for every $\alpha < \mathtt{ot}(\leq_1)$.

We fix an ordinal $\alpha < \mathtt{ot}(\leq_1)$. We assume that for every $\beta < \alpha$, $[\leq_2]_\beta$ is defined and $[\leq_2]_\beta \leq_2 [\leq_1]_\beta$. Our goal is to show that $[\leq_2]_\alpha$ is defined and $[\leq_2]_\alpha \leq_2 [\leq_1]_\alpha$.

Observe that $[\leq_1]_\alpha$ exists, and for every $\beta < \alpha$, $[\leq_2]_\beta \leq_2 [\leq_1]_\beta \leq_2 [\leq_1]_\alpha$. ($[\leq_1]_\beta \leq_1 [\leq_1]_\alpha \rightarrow [\leq_1]_\beta \leq_2 [\leq_1]_\alpha$ because $\leq_1 \subseteq \leq_2$). This means that there is at least one object which is $\geq_2$ all the $[\leq_2]_\beta$'s for $\beta < \alpha$, so there must be a $\leq_2$-least such object x. We claim that $x = [\leq_2]_\alpha$. The objects $\leq_2 x$ are precisely the $[\leq_2]_\beta$'s for $\beta < \alpha$, so the order types of the initial segments of $(\leq_2)_x$ are precisely the ordinals less than α, so the ordinals less than the order type of $(\leq_2)_x$ are precisely the ordinals less than α, and so its order type ... is α as desired.

Now we develop a construction analogous to recursive definition of functions of the natural numbers. Just as transfinite induction is analogous to strong induction on the natural numbers, so transfinite recursion is analogous to course-of-values recursion on the natural numbers.

Transfinite Recursion Theorem: We give a nonce definition of $\mathcal{F}$ as the set of all functions whose domains are segments of the natural order on the ordinals [or on the ordinals less than a fixed γ]:

$$\mathcal{F} = \{f \mid (\exists \alpha \in \mathtt{Ord}. f : \mathtt{seg}_{\leq_\Omega}(\alpha) \rightarrow V)\}.$$

Let G be a function from $\mathcal{F}$ to $\iota``V$. Then there is a unique function g with domain $\mathtt{Ord}$ [or with domain the set of ordinals less than γ] with the property that for every ordinal α [or for every ordinal $\alpha < \gamma$], $\{g(\alpha)\} = G(g\lceil\{\beta \mid \beta < \alpha\})$.

Proof: We say that a set I is G-inductive iff whenever a function $f \in \mathcal{F}$ with domain $\{\beta \in \mathtt{Ord} \mid \beta < \alpha\}$ is a subset of I, $\{\alpha\} \times G(f)$ will be a subset of I. Our claim is that g, defined as the intersection of all G-inductive sets, is the desired function.

We first observe that $\mathtt{Ord} \times V$ is G-inductive, so every element of g actually is an ordered pair whose first projection is an ordinal, as we would expect.

We then prove by transfinite induction on α that $g_\alpha = g \cap \mathtt{seg}_{\leq\Omega}(\alpha) \times V$ is a function with domain $\mathtt{seg}_{\leq\Omega}(\alpha)$. For $\alpha = 0$ this is obvious (the empty set is a function with domain the empty set of all ordinals less than 0). Suppose that g_β is a function with domain the set of ordinals less than β: our goal is then to show that $g_{\beta+1}$ is a function with domain the set of ordinals less than $\beta + 1$. We claim that $X_\beta = g_\beta \cup (\{\beta\} \times G(g_\beta)) \cup (\{\gamma \mid \gamma > \beta\} \times V)$ is G-inductive. Suppose that f is a function with domain the set of ordinals less than δ and f is a subset of X_β. If $\delta < \beta$, it follows that f is a subset of g_β and so $\{\delta\} \times G(f)$ is a subset of g (because g is G-inductive) and also a subset of g_β and so of X_β because the first projection of its sole element is $\delta < \beta$. If $\delta = \beta$, then $f = g_\beta$ and $\{\beta\} \times G(g_\beta)$ is a subset of X_β by construction. If $\delta > \beta$, then $G(f)$ is a subset of X_β because the first projection of its sole element is $\delta > \beta$. From this we can see that $g_\beta \cup G(g_\beta)$ is precisely $g_{\beta+1}$: G-inductiveness of g shows that $g_\beta \cup (\{\beta\} \times G(g_\beta))$ must be included in g, because g_β is included in g; G-inductiveness of X_β shows that g, and so $g_{\beta+1}$, does not include any ordered pairs with first component $\beta + 1$ and second component outside of $G(g_\beta)$. Clearly $g_{\beta+1}$ is a function, with the same value as g_β at each ordinal $< \beta$ and the sole element of $G(g_\beta)$ as its value at β, so its domain is the set of all ordinals less than $\beta + 1$ as desired. Now we consider the case of a limit ordinal λ with the property that g_β is a function for each $\beta < \lambda$. In this case g_λ is the union of all the g_β's. The only way it could fail to be a function is if some two g_β's had distinct values at some ordinal. But this is impossible: it is clear from the definition that $g_\beta \subseteq g_{\beta'}$ for $\beta < \beta'$. It is also obvious that the domain of g_λ is the union of the domains of the g_β's, and the union of the segments determined by the ordinals less than a limit ordinal is the segment determined by that limit ordinal.

Since g is a relation with domain the set of ordinals and its restriction to any initial segment of the ordinals is a function, it is a function. We showed above that the value of $g_{\beta+1}$ (g restricted to the ordinals less than $\beta + 1$) at β is the sole element of $G(g_\beta)$, the value of G at the restriction of g to the ordinals less than β, and this is the recurrence

relation we needed to show. Suppose that $g \neq g'$ were two distinct functions satisfying this recurrence relation. Let δ be the smallest ordinal such that $g(\delta) \neq g'(\delta)$. Note that $\{g(\delta)\} = G(g\lceil\{\gamma \mid \gamma < \delta\}) = G(g'\lceil\{\gamma \mid \gamma < \delta\}) = \{g'(\delta)\}$ by the shared recurrence relation and the fact that g and g' agree at ordinals less than δ, a contradiction.

We give the qualifications needed for a bounded formulation of recursion in brackets in the statement of the theorem: this is the form which would be used in untyped set theory but also in many applications in typed set theory.

We present a variation of the Recursion Theorem:

Transfinite Recursion Theorem: Suppose we are given a set a, a function f and a singleton-valued function F (of appropriate types which can be deduced from the conclusion): then there is a uniquely determined function $g : \mathtt{Ord} \to V$ such that $g(0) = a$, $g(\alpha + 1) = f(g(\alpha))$ for each α, and $g(\lambda)$ is the sole element of $F(\{g(\beta) \mid \beta < \lambda\})$ for each limit ordinal λ.

Proof: This is a special case of the theorem above. The function $G : \mathcal{F} \to \iota\text{“}V$ is defined so that $G(\emptyset) = \{a\}; G(k) = \{f(k(\alpha))\}$ if α is the maximum element of the domain of k; $G(k) = F(\{k(\beta) \mid \beta < \lambda\})$ if the limit ordinal λ is the supremum of the domain of k. The stated recurrence relations are then equivalent to $\{g(\alpha)\} = G(g\lceil\{\beta \mid \beta < \alpha\})$.

The alternative theorem could also be stated in a bounded form.

We define ordinal iteration in a special case. Suppose f is a function and $\leq$ is an order on elements of its field understood from context. Define $f^0(x)$ as x, $f^{\alpha+1}(x)$ as $f(f^\alpha(x))$, and $f^\lambda(x)$ as $\sup\{f^\beta(x) \mid \beta < \lambda\}$. This will uniquely determine a function by either of the recursion theorems. It would seem most natural to do this construction when f was an increasing function in $\leq$ with the property $x \leq f(x)$. A common choice of $\leq$ would be the subset relation.

The arithmetic operations on the ordinals defined above can also be defined by transfinite recursion.

recursive definition of addition: This resembles the iterative definition of addition on the natural numbers.

1. $\alpha + 0 = \alpha$
2. $\alpha + (\beta + 1) = (\alpha + \beta) + 1$
3. $\alpha + \lambda = \sup(\{(\alpha + \beta) \mid \beta < \lambda\})$ when λ is limit.

recursive definition of multiplication: This resembles the iterative definition of multiplication on the natural numbers.

1. $\alpha \cdot 0 = 0$
2. $\alpha \cdot (\beta + 1) = \alpha \cdot \beta + \alpha$
3. $\alpha \cdot \lambda = \sup(\{\alpha \cdot \beta \mid \beta < \lambda\})$ when λ is limit.

recursive definition of exponentiation: Of course a similar definition of exponentiation on natural numbers could be given (and is actually in effect included here). There is a set theoretical definition of exponentiation of ordinals as well, but it is a bit technical.

1. $\alpha^0 = 1$
2. $\alpha^{\beta+1} = \alpha^\beta \cdot \alpha$
3. $\alpha^\lambda = \sup(\{\alpha^\beta \mid \beta < \lambda\})$ when λ is limit.

All the ordinal arithmetic operations commute with the T operation:

Theorem: For any ordinals α and β, $T(\alpha + \beta) = T(\alpha) + T(\beta); T(\alpha \cdot \beta) = T(\alpha) \cdot T(\beta); T(\alpha^\beta) = T(\alpha)^{T(\beta)}$. $T(\alpha) \leq T(\beta) \leftrightarrow \alpha \leq \beta$; if $T^{-\mathbf{1}}(\alpha)$ exists and $T^{-\mathbf{1}}(\beta)$ does not, then $\alpha < \beta$.

We now consider the *original* application of set theory due to Cantor, which includes an example of construction of a function by transfinite recursion. This involves a further discussion of sets of reals.

accumulation point: If X is a set of reals and r is a real number, we say that r is an *accumulation point* of X iff every open interval which contains r contains infinitely many points of X. Note that r does not have to be an element of X to be an accumulation point of X.

closed set: A set of reals X is said to be *closed* iff every accumulation point of X is an element of X.

derived set: For any set X of reals, we define the derived set X' of X as the set of accumulation points of X.

Observations: Obviously X is closed iff $X' \subseteq X$. Whether X is closed or not, X' is closed: if any interval containing r contains infinitely many points of X', then it contains at least one element of X' (accumulation point of X) because it contains infinitely many, and so it contains infinitely many points of X, and so r is itself an accumulation point of X and thus an element of X'. This means further that if we iterate applications of the derived set operator, the first iteration may make our set larger but all subsequent iterations will fix it or remove elements from it.

iteration of the derived set construction: This is a definition by transfinite recursion. Define Δ^X_0 as X. Define $\Delta^X_{\beta+1}$ as $(\Delta^X_\beta)'$. At limit stages, take intersections: define Δ^X_λ as $\bigcap\{\Delta^X_\gamma \mid \gamma < \lambda\}$ for each limit ordinal λ.

Theorem: For every countable ordinal α, there is a set of reals $A \subseteq (0,1]$ with the property that $\Delta^A_\alpha = \{1\}$ (and so $\Delta^A_{\alpha+1} = \emptyset$).

Proof: We prove this by transfinite induction on α. If $\alpha = 0$, the set $\{1\}$ has the desired properties. Suppose that we have a set $A \subseteq (0,1]$ such that $\Delta^A_\beta = \{1\}$. Let f be the function which sends each natural number n to the set A: $f^* \cup \{1\}$ will have the desired property. This set consists of infinitely many successively smaller copies of A approaching the limit point $\{1\}$. Application of the derived set operator β times will reduce each of the infinitely many scale copies of A in $f^* \cup \{1\}$ to a single point.

The next application of the derived set operator will leave just $\{1\}$ (1 is the only accumulation point). So $f^* \cup \{1\}$ is the desired set for which $\beta + 1$ applications of the derived set operator yields $\{1\}$. Now let λ be a countable limit ordinal. There will be a strictly increasing sequence λ_i of ordinals such that λ is the least ordinal greater than all the λ_i's (this is proved above). By inductive hypothesis, we may assume that for each i we have a set A_i such that $\Delta_{\lambda_i}^{A_i} = \{1\}$. Define $f(i) = A_i$ (you might note that this actually requires the Axiom of Choice!). Define $A = f^* \cup \{1\}$. Observe that application of the derived set operator to A $\lambda_i + 1$ times eliminates the copy of A_i, for each i. Notice that application of the derived set operator λ_i times always leaves $\{1\}$ in the set, as the scaled copies of A_j for $j > i$ still have nonempty image, so clearly 1 will still be an accumulation point. It follows from these two observations that the intersection of all the sets $\Delta_{\lambda_i}^{A}$, which will be Δ_{λ}^{A}, will contain no element of any of the original scaled copies of the A_i's but will contain 1: it will be $\{1\}$ as required.

The sets shown to exist by this Theorem are in a sense "discrete" (they cannot be dense in any interval, or no iteration of the derived set operation could eliminate them), but have progressively more complex limit structure calibrated by the countable ordinal α. The applications of these concepts by Cantor to problems in the convergence of trignonometric series are the original motivation (or one of the original motivations) for the development of transfinite ordinals and of set theory.

2.14.1 Exercises

1. Prove that for any ordinals α, β, γ, if $\alpha + \beta = \alpha + \gamma$ then $\beta = \gamma$.'

   You can probably prove this by transfinite induction, using the recursive definitions, but it can be proved using the set theoretic definition and structural properties of ordinals as well.

   Give a counterexample to "if $\beta + \alpha = \gamma + \alpha$ then $\beta = \gamma$".

2. In type theory, prove that for all ordinals α and β, if $\alpha + 1 = \beta + 1$ then $\alpha = \beta$. This is best proved by considering actual well-orderings and isomorphisms between them (not by transfinite induction).

3. Prove by transfinite induction: Every infinite ordinal can be expressed in the form $\lambda + n$, where λ is a limit ordinal and n is a finite ordinal, and moreover it can be expressed in this form in only one way (for this last part you might want to use the result of the previous problem).

2.15 Lateral Functions and T operations; Type-Free Isomorphism Classes

We have observed that cardinals κ and $T^{\mathbf{n}}(\kappa)$, though of different types, are in some sense the same cardinal, and similarly that ordinals α and $T^{\mathbf{n}}(\alpha)$, though of different types, are in some sense the same order type.

We have $T^{\mathbf{n}}(|A|) = |B|$ iff $|\iota^{\mathbf{n}}\text{“}A| = |B|$, that is iff there is a bijection $f : \iota^{\mathbf{n}}\text{“}A \to B$. The bijection f witnesses the fact that A and B are "the same size", by exploiting the fact that A and $\iota^{\mathbf{n}}\text{“}A$ are externally "the same size".

We introduce the following definitions.

Definition (lateral relations): If $R \subseteq \iota^{\mathbf{n}}\text{“}A \times B$, we define $x\, R_{\mathbf{n}}\, y$ as holding iff $\iota^{\mathbf{n}}(x)\, R\, y$. Similarly, if $S \subseteq A \times \iota^{\mathbf{n}}\text{“}B$, we define $x\, S_{-\mathbf{n}}\, y$ as holding iff $x\, S\, \iota^{\mathbf{n}}(y)$.

Definition (description of lateral relations): We define $A \times_{\mathbf{n}} B$ as $\iota^{\mathbf{n}}\text{“}A \times B$ and
$A \times_{-\mathbf{n}} B$ as $A \times \iota^{\mathbf{n}}\text{“}B$.

Definition (lateral functions): If $f : \iota^{\mathbf{n}}\text{“}A \to B$, we define $f_{\mathbf{n}}(a) = f(\iota^{\mathbf{n}}(a))$ for each $a \in A$. Similarly, if $g : A \to \iota^{\mathbf{n}}\text{“}B$, we define $g_{-\mathbf{n}}(a) = \iota^{-\mathbf{n}}(g(a))$.

Definition (description of lateral functions): $f_{\mathbf{n}} : A \to B$ is defined as $f : \iota^{\mathbf{n}}\text{“}A \to B$; $f_{-\mathbf{n}} : A \to B$ is defined as $f : A \to \iota^{\mathbf{n}}\text{“}B$.

Note that in none of these notations is a boldface subscript actually part of the name of a function or relation: the boldface subscripts are always indications of the role the function or relation is playing in the expression.

This definition allows us to code relations and functions with domains and ranges of different types. Note that this definition allows us to say that $T^{\mathbf{n}}(|A|) = |B|$ iff there actually is a (lateral) bijection from A to B! The definition also allows us to assert that well-orderings of types α and $T^{\mathbf{n}}(\alpha)$ actually are "isomorphic" in the sense that there is a lateral function satisfying the formal conditions to be an isomorphism between them.

We present the Transfinite Recursion Theorem in a slightly different format:

Transfinite Recursion Theorem: We give a nonce definition of $\mathcal{F}$ as the set of all functions whose domains are segments of the natural order on the ordinals [or on the ordinals less than a fixed γ]. Let $G_{-\mathbf{1}} : \mathcal{F} \to \mathcal{V}$. Then there is a unique function g with domain $\mathtt{Ord}$ [or with domain the set of ordinals less than γ] with the property that for every ordinal α [or for every ordinal $\alpha < \gamma$], $g(\alpha) = G_{-\mathbf{1}}(g\lceil \mathtt{seg}(\alpha))$.

We give a general "comprehension" theorem for functions and relations with a type differential.

Theorem: If $\phi[x^n, y^{n+k}]$ is a formula, there is a set relation R such that $x\, R_{\mathbf{k}}\, y \leftrightarrow \phi[x, y]$ (where types revert to being implicit in the second formula).

If $\phi[x^{n+k}, y^n]$ is a formula, there is a set relation R such that $x\, R_{-\mathbf{k}}\, y \leftrightarrow \phi[x, y]$ (where types revert to being implicit in the second formula).

If $(\forall x^n \in A.(\exists! y^{n+k}.\phi[x^n, y^{n+k}]))$, then there is a function $f_{\mathbf{k}} : A \to V$ such that for any $x \in A$, $y = f_{\mathbf{k}}(x) \leftrightarrow \phi[x, y]$.

If $(\forall x^{n+k} \in A.(\exists! y^n.\phi[x^{n+k}, y^n]))$, then there is a function $f_{-\mathbf{k}} : A \to V$ such that for any $x \in A$, $y = f_{-\mathbf{k}}(x) \leftrightarrow \phi[x, y]$.

Corollary: If A^n and B^{n+k} are sets and there is a formula $\phi[a, b]$ such that $(\forall a \in A.(\exists! b \in B.\phi[a, b])) \wedge (\forall b \in B.(\exists! a \in A, \phi[a, b]))$, then $T^k(|A|) = |B|$. If $\leq_1^n$ and $\leq_2^{n+k}$ are well-orderings, and there is a formula ϕ such that $(\forall xy.x <_1 y \leftrightarrow (\exists zw.\phi[x, z] \wedge \phi[y, w] \wedge z <_2 w)) \wedge (\forall zw.z <_2 w \leftrightarrow (\exists xy.\phi[z, x] \wedge \phi[w, y] \wedge x <_1 y))$, then $T^k(\mathtt{ot}(\leq_1)) = \mathtt{ot}(\leq_2)$.

All parts of this theorem are proved by direct application of the Axiom of Comprehension. The Corollary expresses the idea that any external bijection or isomorphism we can describe using a formula is actually codable by a set and so witnesses appropriate cardinal or ordinal equivalences.

We note that T operations can be defined for general isomorphism classes.

Definition: For any relation R, the isomorphism class $[R]_{\approx} = \{S \mid R \approx S\}$. We define $T([R]_{\approx}) = [R^\iota]_{\approx}$, where $R^\iota = \{\langle \{x\}, \{y\}\rangle \mid x\, R\, y\}$, as already defined. Note that this is more general than but essentially the same as the T operation on ordinals.

Now we pursue an extension of the Reasonable Convention proposed above for natural numbers. We recall that the T operation on cardinals witnesses an exact correspondence between the natural numbers at different types. This allows us, if we wish, to introduce natural number variables which can be used in a type-free manner: such a variable can be shifted into the type appropriate for any context by appropriate applications of the T operation or its inverse. All statements purely in the language of the natural numbers are invariant under uniform application of the T operation, as we have seen. Each occurrence of a natural number variable translates into an occurrence of a general variable of an appropriate type restricted to the set of natural numbers at the appropriate type.

This idea can be extended to cardinals and ordinals (and to isomorphism classes in general), but a further refinement is needed. The difficulty is that the ordinals in one type are mapped injectively into but not onto the ordinals in the next type, as we have just seen. We will see below that the same is true of the cardinals. The natural number variables introduced in the previous paragraph are translated as general variables restricted to the set of all natural numbers (which is in effect the same set at each type); this cannot work for the ordinals (or the cardinals): each ordinal bound variable must be restricted to the ordinals in a specific type (which is equivalent to restriction to an initial segment of "all the ordinals" determined by the first ordinal not in that particular type (the first ordinal of the next higher type which is not an image under T)). We can thus use type-free ordinal variables as long as we require that any such variable be restricted to a proper initial segment of the ordinals (the type of the bound will determine the highest type in which we can be working), and we can treat cardinals similarly. There is no way to express a general assertion about all ordinals at whatever type in type theory. Just as in natural number arithmetic, all statements about properties, relations, and operations natural to cardinals and ordinals are invariant under uniform application of the T operation: this enables the proposed identifications of cardinals and ordinals at diverse types to cohere.

This convention would allow the elimination in practice of the inconvenient reduplication of cardinals, ordinals, and similar constructions at each type. We do not use it as yet, but it is important to us to note that it is possible to use this convention.

2.15.1 Lateral functions in the system of the unsorted preamble: an axiom of embedding

In the system of section 2.1.1, where we cannot be sure that all types are indexed by natural numbers as in our basic type scheme, we postulate the existence of many external embeddings of one type into another.

Axiom of embedding: We postulate a primitive strict partial order $<_\tau$ on types. We postulate function symbols I_{xy} and I_{yx} such that we have $I_{xy}(a)$ defined just in case $a \in \tau(x)$, $I_{xy}(a) \in \tau(y)$ if it is defined, and, if $\tau(x) <_\tau \tau(y)$, $I_{yx}(I_{xy}(a)) = a$ for all $a \in \tau(x)$. We provide further that for any x, y there is z such that $\tau(x) <_\tau (z)$ and $\tau(y) <_\tau (z)$ and that further, $\tau(x) <_\tau (y)$ iff $|I_{xz}\text{“}(\tau(x))| < |I_{yz}\text{“}(\tau(y))|$.

Further clauses which are tempting to add to our axiom, though we believe they are not really needed, are $I_{x\tau(x)}(a) = \{a\}$ and, for $\tau(x) <_\tau (y) <_\tau (z)$, $I_{yz}(I_{xy}(a)) = I_{xz}(a)$ for all $a \in \tau(x)$. Another reasonable assumption is that the action of the function symbol I_{xy} is exactly determined by the type of x and the type of y.

I want to include a description here of how operations such as $\iota^n(x)$ can actually be defined in terms of the natural numbers of the theory using this axiom.

An effect of this axiom is that isomorphism types (including cardinal and ordinal numbers) can be systematically identified between types.

This axiom implies the existence of base types in the unsorted system.

2.16 Other Forms of the Axiom of Choice

The Axiom of Choice is equivalent to some other interesting propositions (in fact, there are whole books of them but we will only discuss a few).

The Well-Ordering Theorem: Every set is the field of a well-ordering. (Equivalently, V is the field of a well-ordering.)

Observation: It is obvious that the well-ordering theorem implies the Axiom of Choice: the choice set of a partition can be taken to be the set of minimal elements in the elements of the partition under a well-ordering of the union of the partition. The interesting part of the result is the converse: the Axiom of Choice is enough to prove the Well-Ordering Theorem.

Definition: A *chain in a partial order* $\leq$ is a subset C of $\mathtt{fld}(\leq)$ such that $\leq \cap C^2$, the restriction of $\leq$ to C, is a linear order (i.e., any two elements of C are comparable in the order).

Definition: A collection of sets is said to be *nested* iff it is a chain in the inclusion order: A is a nested collection of sets iff $(\forall x \in A.(\forall y \in A.x \subseteq y \vee y \subseteq x))$.

Lemma: The union of a nested collection of chains in a partial order $\leq$ is a chain in $\leq$.

Zorn's Lemma: A partial order with nonempty domain in which every chain has an upper bound has a maximal element.

Observation: Let $\mathcal{A}$ be the set of all well-orderings of subsets of a set A. We define $U \leq V$ as holding for $U, V \in \mathcal{A}$ iff either $U = V$ or U is a segment restriction of V. A chain in this well-ordering is a collection C of well-orderings of A which agree with one another in a strong sense and whose union will also be a well-ordering of a subset of A and so an upper bound of the chain C (details of this bit are left as an exercise). So Zorn's Lemma would allow us to conclude that there was a maximal partial well-ordering of A under the segment restriction relation, which clearly must be a well-ordering of all of A (any element not in the field of the maximal well-order could be adjoined as a new largest element of a larger well-ordering for a contradiction).

Since Zorn implies Well-Ordering and Well-Ordering implies Choice, it only remains to show that Choice implies Zorn to prove that all three are equivalent (in the presence of the rest of our axioms).

Proof of Zorn's Lemma: Let $\leq$ be a partial order in which every chain has an upper bound.

Let $\mathcal{C}$ be the set of all chains in $\leq$. Note that for any chain C if there is an upper bound of C which belongs to C there is exactly one such upper bound. If in addition all upper bounds of C belong to C then this uniquely determined upper bound is maximal in $\leq$. For each chain C in $\leq$, define B_C as the set of all upper bounds for C which are not in C, if there are any, and otherwise as the singleton of the unique upper bound of C which is an element of C. All of these sets will be nonempty. The set $\{\{\langle C, \{b\}\rangle \mid b \in B_C\} \mid C \in \mathcal{C}\}$ is a partition, and so has a choice set. Notice that the choice set is a function F which sends each $C \in \mathcal{C}$ to the singleton set of an upper bound of C, which will belong to C only if all upper bounds of C belong to C (in which case the upper bound is maximal).

For each chain C, denote the linear order $\leq \cap C^2$ by $\leq_C$. We call a chain C a *special chain* iff $\leq_C$ is a well-ordering and for each $x \in C$ we have $\{x\} = F(\mathtt{fld}((\leq_C)_x))$.

We can prove by transfinite induction that $\leq_C$ is precisely determined by its order type (for any special chains C and D, if $\leq_C$ is isomorphic to $\leq_D$ then $\leq_C = \leq_D$). Suppose otherwise: then there is a least ordinal to which distinct $\leq_C$ and $\leq_D$ belong. There must be a $\leq_C$-first element x which differs from the corresponding $\leq_D$ element y. But this implies that $(\leq_C)_x = (\leq_D)_y$ whence $\{x\} = F((\leq_C)_x) = F((\leq_D)_y) = \{y\}$.

This implies further that for any two distinct special chains, one is a segment restriction of the other. This further implies that the union of all special chains is a linear order and in fact a special chain; call it E. Now $E \cup F(\leq_E)$ is a special chain as well, which cannot properly extend E, so $F(\leq_E) \subseteq E$, so the sole element of $F(E)$ is a maximal element with respect to $\leq$.

Alternative Proof of Zorn's Lemma: Let $\leq$ be a nonempty partial order in which any chain has an upper bound. Let $\mathcal{C}$ be the set of all chains in $\leq$.

For each chain C in $\leq$ and $x \in \mathtt{fld}(\leq)$, we say that x is an appropriate upper bound of C if x is an upper bound of C and $x \notin C$ or if $x \in C$ and all upper bounds of C are elements of C. Notice that if there is an upper bound of C belonging to C there is only one, and also notice that if the unique upper bound of C belonging to C is the only upper bound of C then it is maximal in $\leq$, because anything strictly greater than the unique upper bound of C in C would be an upper bound of C not in C.

For each chain C in $\leq$, we define X_C as the set of all ordered pairs $\langle C, \{x\}\rangle$ such that x is an appropriate upper bound of C. Notice that if $C \neq C'$ then X_C and $X_{C'}$ are disjoint (because elements of the two sets are ordered pairs with distinct first projections). Thus $\{X_C \mid C \in \mathcal{C}\}$ is a partition, and has a choice set F. Notice that F is a function, $F : \mathcal{C} \to \iota\text{“}\mathtt{fld}(\leq)$, and $F(C)$ for every C is the singleton $\{x\}$ of an appropriate upper bound x of C.

Define a function G by transfinite recursion: $G(\alpha)$ is defined as the sole element of $F(\mathtt{rng}(G\lceil\{\beta \mid \beta < \alpha\}))$ if $\mathtt{rng}(G\lceil\{\beta \mid \beta < \alpha\})$ is a chain in $\leq$ and as 0 otherwise. Transfinite induction shows that $\mathtt{rng}(G\lceil\{\beta \mid \beta < \alpha\})$ is a chain in $\leq$ for any ordinal α (for successor α, because $C \cup F(C)$ is always a chain in $\leq$ if C is a chain in $\leq$, and for limit α because a union of nested chains in $\leq$ is a chain in $\leq$). Note that $G(\alpha)$ will not be one of the $G(\beta)$'s for $\beta < \alpha$ unless it is maximal for $\leq$, so G is injective if $\leq$ has no maximal element. The range of G is a subset of the field of $\leq$ with a unique well-ordering under which G is an increasing function. The order type of this well-ordering will be the order type Ω of the ordinals iff G is injective. If G is not injective, it is constant past a certain point and so the order type of this well-ordering will be that of an initial segment of the ordinals, so strictly less than Ω. Now we employ a trick: consider instead of $\leq$ the order type $\leq^{\iota^{\mathbf{2}}}$ of double singletons induced by $\leq$. The well-ordering of the range of the function G associated with $\leq^{\iota^{\mathbf{2}}}$ will have some order type $T^{\mathbf{2}}(\alpha) < \Omega$ (because it is a well-ordering of a set of double singletons) and so cannot be injective, and so $\leq^{\iota^{\mathbf{2}}}$ has a maximal element, from which it follows that $\leq$ itself has a maximal element. The point of the

trick is that the original working type we started with might not have had enough ordinals for the construction of G to exhaust the field of $\leq$.

Throughout this discussion we could have used the lateral function notation introduced in the previous subsection: $F_{-\mathbf{1}}(C)$ is an upper bound for C for each chain C.

NOTE: include examples of use of Zorn's Lemma in other parts of mathematics.

The Axiom of Choice directly enables us to make choices from pairwise disjoint collections of sets. But in fact we can use the Axiom to show that we can make choices from any collection of nonempty sets.

Definition: Let A be a collection of nonempty sets. A function c with domain A is called a *choice function for* A iff $c : A \to 1$ ($c(a)$ is a one element set for each $a \in A$) and $c(a) \subseteq a$ for each $a \in A$. The sole element of $c(a)$ is the item selected from A by the choice function.

It is equivalent to say (using the notation for lateral functions) that a choice function for A is a function $c_{-\mathbf{1}} : A \to V$ such that $c_{-\mathbf{1}}(a) \in a$ for each $a \in A$.

Theorem: Each collection of nonempty sets A has a choice function.

Proof: The collection $\{\{a\} \times \iota\text{“}a \mid a \in A\}$ is a partition and so has a choice set c. This choice set is the desired choice function.

We observe that a logical device proposed above and revisited from time to time, but not officially adopted, can be introduced by definition at this point.

Hilbert symbol: Let H be a fixed function $V \to 1$ such that $H\lceil(\mathcal{P}(V) - \{\emptyset\})$ is a choice function and $H(\emptyset) = \emptyset$ (if the type of the second $\emptyset$ is positive). We do not care which one. Define $(\epsilon x.\phi)$ as the sole element of $H(\{x \mid \phi\})$ for each formula ϕ.

Theorem: For any formula ϕ, $(\exists x.\phi) \leftrightarrow \phi[(\epsilon x.\phi)/x]$. Since $(\forall x.\phi) \leftrightarrow \neg(\exists x.\neg\phi)$, this means that both quantifiers could be defined in terms of the Hilbert symbol.

Proof: This is obvious.

Note that a systematic use of the Hilbert symbol would imply a choice of an H in each relevant type.

2.16.1 Exercises

1. Prove that the union of a nested set of chains in a partial order $\leq$ is a chain. A chain is a set C such that for any $x, y \in C$ we have either $x \leq y$ or $y \leq x$; a nested collection of sets is a set A of sets which is a chain in the subset relation (for any $x, y \in A$, either $x \subseteq y$ or $y \subseteq x$).

2. Prove that the union of a countably infinite collection of countably infinite sets is countably infinite. Notice that you already know that $\mathbb{N} \times \mathbb{N}$ is a countable set.

   We give the result in more detail: suppose that F is a function with domain $\mathbb{N}$ and the property that each $F(n)$ is a countably infinite set. Show that $\bigcup\{F(n) \mid n \in \mathbb{N}\}$ is countable (that is, show that it is the range of a bijection with domain the set of natural numbers).

   Hint: be very careful. It is fairly easy to see why this is true if you understand why $\mathbb{N} \times \mathbb{N}$ is a ocuntable set, but there is an application of the Axiom of Choice involved which you need to notice; in type theory or set theory without choice there may be countable collections of countable sets which have uncountable unions!

3. Use Zorn's Lemma to prove that every infinite set is the union of a pairwise disjoint collection of countably infinite sets.

   Then prove that if B is a collection of countably infinite sets, $|\bigcup B| = |\bigcup B| + |\bigcup B|$. (This exploits the fact that $|\mathbb{N}| = |\mathbb{N}| + |\mathbb{N}|$; it also requires the Axiom of Choice).

   Notice that this is another proof that $\kappa + \kappa = \kappa$ for any infinite cardinal κ.

2.17 Transfinite Arithmetic of Order, Addition, and Multiplication

We define the order relation on cardinals in a natural way.

order on cardinals: $|A| \leq |B|$ iff there is an injection from A to B.

Implicit in our notation is the claim that $\leq$ is a partial order. The relation is obviously reflexive and transitive: that it is antisymmetric is a famous theorem.

Cantor-Schröder-Bernstein Theorem: If $|A| \leq |B|$ and $|B| \leq |A|$ then $|A| = |B|$.

Before proving this theorem we give an example to illustrate why it is not obvious. Consider the sets $[0,1] = \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$ and $\mathcal{P}(\mathbb{N})$, the set of all sets of natural numbers.

A injection f from $[0,1]$ into $\mathcal{P}(\mathbb{N})$ is defined by $f(r) = \{k \in \mathbb{N} \mid \frac{1}{2^{k+1}}$ is a term in the unique nonterminating binary expansion of $r\}$ while a bijection g from $\mathcal{P}(\mathbb{N})$ into $[0,1]$ is given by $g(A) = \Sigma_{k \in A} \frac{1}{10^{k+1}}$. So it is easy to see that each set embeds injectively in the other, but it is not at all easy to see how to construct a bijection which takes one set exactly onto the other.

We now give the slightly delayed

Proof of the Cantor-Schröder-Bernstein Theorem: Assume that there is an injection $f : A \to B$ and an injection $g : B \to A$: our goal is to show that there is a bijection h from A to B. B is the same size as $g\text{“}B \subseteq A$, so if we can show $A \sim g\text{“}B$ we are done. The map $f|g$ sends all of A into $g\text{“}B$; we develop a trick to send it exactly onto $g\text{“}B$. Let C be the intersection of all sets which contain $A - g\text{“}B$ and are closed under $f|g$. Let h_0 be the map which sends all elements of C to their images under $f|g$ and fixes all elements of $A - C$. This is a bijection from A to $g\text{“}B$, so $h_0|g^{-1}$ is a bijection from A to B.

Note that this proof does not use the Axiom of Choice. Beyond this point we will use the Axiom of Choice freely, and some of the results we state are not necessarily true in type theory or set theory without Choice.

Theorem: The natural order on cardinals is a linear order.

Proof: Let A and B be sets: we want to show $|A| \leq |B|$ or $|B| \leq |A|$. This is easy using the Well-Ordering Theorem: we choose well-orderings $\leq_A$ and $\leq_B$ of A and B respectively. If the well-orderings are isomorphic, the isomorphism between them witnesses $|A| = |B|$ (and so $|A| \leq |B|$. Otherwise, one of $\leq_A$ and $\leq_B$ is isomorphic to a segment restriction of the other, and the isomorphism is the required injection from one of the sets into the other.

Theorem: The natural order on cardinals is a well-ordering.

Proof: Let C be a set of cardinals. Our aim is to show that C has a smallest element in the natural order. Let $\leq$ be a well-ordering of a set at least as large as any of the elements of the union of C (the universe of the appropriate type will work). Consider the set of all well-orderings of elements of the union of C (note that the union of C is the set of all sets which have cardinalities in the set C). Every well-ordering in this set will either be similar to $\leq$ or similar to some segment restriction of $\leq$. If all are similar to $\leq$, then all elements of C are the same and it has a smallest element. Otherwise consider the set of all x such that $\leq_x$ is isomorphic to some well-ordering of some element of the union of C: there must be a $\leq$-smallest element of this set, which corresponds to the smallest element of C in the natural order.

Theorem: There is a surjection from A onto B iff $|B| \leq |A|$ (and B is nonempty if A is).

Proof: If there is an injection f from B to A, then we can define a surjection from A to B as follows: choose $b \in B$; map each element of A to $f^{-1}(a)$ if this exists and to b otherwise. This will be a surjection. If B is empty we cannot choose b, but in this case A is empty and there is obviously a surjection.

If there is a surjection f from A onto B, there is a partition of A consisting of all the sets f"$\{a\}$ for $a \in A$. Let C be a choice set for this partition. Map each element b of B to the unique element of $C \subseteq A$ which is sent to b by f. This map is obviously an injection.

Definition: In type theory or set theory *without* Choice, we define

$$|A| \leq^* |B|$$

as holding iff there is a surjection from B onto A. In the light of the previous Theorem, there is no need for this notation if we assume Choice.

Theorem: For all cardinals κ and λ, $\kappa \leq \lambda \leftrightarrow T(\kappa) \leq T(\lambda)$. If $T^{-\mathbf{1}}(\kappa)$ exists and $T^{-\mathbf{1}}(\lambda)$ does not exist then $\kappa \leq \lambda$.

Definition (repeated from above): We define $\aleph_0$ as $|\mathbb{N}|$. Elements of $\aleph_0$ are called *countably infinite sets*, or simply *countable sets*.

Theorem: $\aleph_0 + 1 = \aleph_0 + \aleph_0 = \aleph_0 \cdot \aleph_0 = \aleph_0$. It is straightforward to define a bijection between $\mathbb{N}$ and $\mathbb{N} \times \mathbb{N}$. The bijections between the naturals and the even and odd numbers witness the second statement. The successor map witnesses the first statement.

Theorem: $\aleph_0 = T(\aleph_0)$.

Proof: This follows from the fact that natural numbers are sent to natural numbers by the T operation and by its inverse.

Theorem: Every infinite set has a countable subset.

Proof: Let A be an infinite set. The inclusion order on the collection of all bijections from initial segments of $\mathbb{N}$ to A satisfies the conditions of Zorn's Lemma and so has a maximal element. If the maximal element had domain a proper initial segment of $\mathbb{N}$, then the set would be finite. So the maximal element is a bijection from $\mathbb{N}$ to a subset of A.

Theorem: For every infinite cardinal κ, $\kappa + 1 = \kappa$.

Proof: Let A be an infinite set. The inclusion order on the set of all bijections from B to $B \cup \{x\}$, where $B \cup \{x\} \subseteq A$ and $x \notin B$, satisfies the conditions of Zorn's Lemma and so has a maximal element. It is nonempty because A has a countable subset. If the maximal element is a map from B to $B \cup \{x\}$ and there is $y \in A - (B \cup \{x\})$, then affixing $\langle y, y \rangle$ to the map shows that the supposed maximal element was not maximal.

An easier proof of this uses the previous theorem. $\kappa = \lambda + \aleph_0$ for some λ, since a set of size κ has a countable subset. It follows that $\kappa = \lambda + \aleph_0 = \lambda + (\aleph_0 + 1) = (\lambda + \aleph_0) + 1 = \kappa + 1$.

Corollary: If n is finite and κ is an infinite cardinal then $\kappa + n = \kappa$.

Theorem: For every infinite cardinal κ, $\kappa + \kappa = \kappa$.

Proof: Let A be an infinite set. The set of pairs of injections f and g with $\mathtt{dom}(f) = \mathtt{dom}(g) = \mathtt{rng}(f) \cup \mathtt{rng}(g) \subseteq A$ and $\mathtt{rng}(f) \cap \mathtt{rng}(g) = \emptyset$ can be partially ordered by componentwise inclusion:

$$(f, g) \leq (f', g') \leftrightarrow f \subseteq f' \wedge g \subseteq g'.$$

This partial order satisfies the hypotheses of Zorn's Lemma (verifying this is left as an exercise). It is nonempty because A has a countable subset. Suppose that a maximal such pair of bijections f, g in the componentwise inclusion order has been constructed. Let B be the common domain of f and g. If there is no countably infinite subset in $A - B$, then $A - B$ is finite and $|B| = |A|$ by a previous result and the result is proved: otherwise take a countable subset of $A - B$ and extend the supposedly maximal pair of maps to a larger one.

Corollary: If $\lambda \leq \kappa$ and κ is an infinite cardinal then $\kappa + \lambda = \kappa$: note that $\kappa \leq \kappa + \lambda \leq \kappa + \kappa = \kappa$.

Theorem: For every infinite cardinal κ, $\kappa \cdot \kappa = \kappa$.

Proof: Let A be an infinite set. The inclusion order on bijections from $B \times B$ to B, where $B \subseteq A$, satisfies the conditions of Zorn's Lemma. It is nonempty because A has a countable subset. Now consider a maximal function in this order, mapping $B \times B$ to B. If $A - B$ contains no subset as large as B, then $|B| = |A|$ by the previous result and the result is proved. Otherwise, choose $B' \subseteq A - B$ with $|B'| = |B|$. It is then easy to see from assumptions about B and B' and the previous result that the map from $B \times B$ to B can be extended to a bijection from $(B \cup B') \times (B \cup B')$ to $B \cup B'$, contradicting the supposed maximality of the bijection.

Corollary: If $\lambda \leq \kappa$ and κ is an infinite cardinal then $\kappa \cdot \lambda = \kappa$.

The arithmetic of addition and multiplication of infinite cardinals is remarkably simple. This simplicity depends strongly on the presence of Choice.

2.17.1 Exercises

1. A classical argument that $|\mathcal{R}^2| = |\mathcal{R}|$ goes as follows. Suppose that it is granted that $|[0,1]| = |\mathcal{R}|$ (this takes a wee bit of work, too, but not too much). So it suffices to prove that $|[0,1]|^2 = |[0,1]|$. Map the pair of numbers with decimal expansions $0.a_1a_2a_3\ldots$ and $0.b_1b_2b_3\ldots$ to the number with expansion $0.a_1b_1a_2b_2a_3b_3\ldots$. Unfortunately, this doesn't quite give us the bijection we want due to problems with decimal expansions (explain). Give a corrected description of this map, taking into account bad features of decimal expansions, and explain why it is not a bijection from $[0,1]^2$ to $[0,1]$. Is it an injection? A surjection? Then use a theorem from the notes (giving all details of its application to this situation) to conclude that there is a bijection from $[0,1]^2$ to $[0,1]$.

2.18 Cantor's Theorem

2.18.1 Cardinal Exponentiation and the Theorem

In this section, we start by defining another arithmetic operation. We have delayed this because its properties in the transfinite context are more vexed.

Definition (function space): The set of all functions from A to B is called B^A. Note that B^A is one type higher than A or B (it would be three types higher if we were using the Kuratowski pair).

Definition (cardinal exponentiation): We define $|A|^{|B|}$ as $T^{-\mathbf{1}}(A^B)$.

The appearance of $T^{-\mathbf{1}}$ is required to get a type-level operation (it would be $T^{-\mathbf{3}}$ if we used the Kuratowski pair). It makes it formally possible that this operation is partial – and indeed it turns out that this operation *is* partial.

Definition: For each subset B of A define χ^A_B as the function from A to $\{0,1\}$ which sends each element of B to 1 and each element of $A - B$ to 0. We call this *the characteristic function of B (relative to A).*

Observation: The function sending each $B \subseteq A$ to χ^A_B is a bijection.

Theorem: $|\mathcal{P}(A)| = |\{0,1\}^A|$, so $2^{|A|} = T^{-\mathbf{1}}(|\mathcal{P}(A)|)$.

Now comes the exciting part.

Cantor's Theorem: For any set A, there is no function f from $\iota``A$ onto $\mathcal{P}(A)$.

Proof: Suppose otherwise, that is, that there is a function f from $\iota``A$ onto $\mathcal{P}(A)$. Consider the set

$$R = \{a \in A \mid a \notin f(\{a\})\}.$$

Since f is onto, $R = f(\{r\})$ for some $r \in A$. Now

$$r \in R \leftrightarrow r \notin f(\{r\}) = R$$

is a contradiction.

This tells us that a set A cannot be the same size as its power set. The fact that A and $\mathcal{P}(A)$ are of different types necessitates the exact form of the theorem. This implies that if $2^{|A|}$ exists, that

$$T(|A|) = |\iota\text{“}A| \neq |\mathcal{P}(A)| = T(2^{|A|})$$

so $|A| \neq 2^{|A|}$. There are at least two distinct infinite cardinals, $|\mathbb{N}|$ and $2^{|\mathbb{N}|}$ (in high enough types for both to be present).

Since certainly $|\iota\text{“}A| \leq |\mathcal{P}(A)|$ (singletons of elements of A are subsets of A), it follows by Cantor-Schroder-Bernstein that $|\mathcal{P}(A)| \not\leq |\iota\text{“}A|$, as otherwise these two cardinals would be equal, so we can write $|\iota\text{“}A| < |\mathcal{P}(A)|$ and $\kappa < 2^\kappa$.

Further we have the curious result that $|\iota\text{“}V| < |\mathcal{P}(V)|$ must be distinct: there are more sets in any given type than singletons of sets (of the next lower type). This implies that V in any given type is strictly larger than any set of lower type (in the sense that the elementwise image under an appropriate $\iota^{\mathbf{n}}$ of the lower type set at the same type as V will have smaller cardinality than V): $T^{-\mathbf{1}}(|V|)$ is undefined and so is $2^{|V|}$, which would be $T^{-\mathbf{1}}(|\mathcal{P}(V)|)$.

2.18.2 Applications to the Number Systems

We give some set theoretical facts about familiar number systems.

Theorem: $\mathbb{Z} \sim \mathbb{N}$

Proof: Consider the map which sends 0 to 0, $2n-1$ to n for each natural number $n > 0$ and $2n$ to $-n$ for each $n > 0$. This is a bijection.

Theorem: $\mathbb{Q} \sim \mathbb{N}$.

Proof: There is an obvious injection from $\mathbb{Q}$ into $\mathbb{Z} \times \mathbb{N}^+$ determined by simplest forms of fractions. $\mathbb{Z} \times \mathbb{N}^+ \sim \mathbb{N} \times \mathbb{N}$ is obvious. $\mathbb{N} \times \mathbb{N}$ is injected into $\mathbb{N}$ by the map $f(m,n) = 2^m 3^n$, and of course $\mathbb{N}$ injects into $\mathbb{Q}$. The result follows by the Cantor-Schröder-Bernstein theorem.

Theorem: $\mathbb{R} \sim \mathcal{P}(\mathbb{N})$, so $|\mathbb{R}| > |\mathbb{N}|$.

Proof: An injection from the interval $[0,1)$ in the reals into $\mathcal{P}(\mathbb{N})$ is defined by sending each real r in that interval to the set of all natural numbers i such that there is a 1 in the $\frac{1}{2^i}$'s place in the binary expansion of

r which contains infinitely many 1's. An injection from $\mathcal{P}(\mathbb{N})$ to the interval $[0,1)$ sends each set A of natural numbers to the real number whose base 3 expansion consists of 1's in the $\frac{1}{3^i}$'s place for each $i \in A$ and zeroes in all other places. It follows by Cantor-Schröder-Bernstein that

$$[0,1) \sim \mathcal{P}(\mathbb{N}).$$

Injections from $(-\frac{\pi}{2}, \frac{\pi}{2})$ into $[0,1)$ and vice versa are easy to define, so

$$(-\frac{\pi}{2}, \frac{\pi}{2}) \sim [0,1).$$

Finally, the arc tangent function witnesses

$$(-\frac{\pi}{2}, \frac{\pi}{2}) \sim \mathbb{R},$$

The cardinal inequality follows from Cantor's Theorem.

The linear orders on $\mathbb{Q}$ and $\mathbb{R}$ share a characteristic which might suggest to the unwary that both sets should be larger than the "discrete" $\mathbb{N}$.

Definition: If $\leq$ is a linear order and $A \subseteq \mathtt{fld}(\leq)$, we say that A is *dense in* $[\leq]$ iff for each $x < y$ there is $z \in A$ such that $x < z$ and $z < y$ (it is traditional to write $x < z \wedge z < y$ as $x < z < y$). We say that $\leq$ itself is merely *dense* iff $\mathtt{fld}(\leq)$ is dense in $[\leq]$.

Observation: The natural orders on $\mathbb{Q}$ and $\mathbb{R}$ are dense. $\mathbb{Q}$ is dense in the order on $\mathbb{R}$.

Definition: A linear order with a finite or countable dense set is said to be *separable*. The immediately preceding example shows that a separable linear order need not be countable.

Theorem: Any two dense linear orders with countably infinite field and no maximum or minimum element are isomorphic. This is a characterization of the order on $\mathbb{Q}$ up to isomorphism.

Proof: Let $\leq_1$ and $\leq_2$ be two such orders. Let $\leq^1$ and $\leq^2$ be well-orderings of order type ω with the same fields as $\leq_1$ and $\leq_2$ respectively.

We define a map f from $\mathtt{fld}(\leq_1)$ to $\mathtt{fld}(\leq_2)$ by a recursive process.

Initially, we define a_1^0 as the $\leq^1$-least element of $\texttt{fld}(\leq_1)$, and define $f(a_1^0)$ as the $\leq^2$-least element of $\texttt{fld}(\leq_2)$. This completes stage 0 of the construction.

Suppose that the values at which f has been defined at the nth stage of our construction are the terms $a_i^n (0 \leq i \leq N)$ of a finite strictly $\leq_1$-increasing sequence of elements of $\texttt{fld}(\leq_1)$, and further that $f(a_i^n)(0 \leq i \leq N)$ is a strictly increasing $\leq_2$- sequence. We define a_{2i+1}^{n+1} as a_i^n for each i in the domain of a^n. Note that this means that f is already defined at each of the odd-indexed elements of the range of a^{n+1} that we will consider in what follows. We define a_0^{n+1} as the $\leq^1$-least element of the $\leq_1$-interval $(-\infty, a_1^{n+1})$ and $f(a_0^{n+1})$ as the $\leq^2$-least element of the $\leq_2$-interval $(-\infty, f(a_1^{n+1}))$. We define a_{2N+2}^{n+1} as the $\leq^1$-least element of the $\leq_1$-interval (a_{2N+1}^{n+1}, ∞), and $f(a_{2N+2}^{n+1})$ as the $\leq^2$-least element of the $\leq_2$-interval $(f(a_{2N+1}^{n+1}), \infty)$. These selections succeed because neither order has a maximum or minimum. For $0 \leq i < N$, we define a_{2i}^{n+1} as the $\leq^1$-least element of the $\leq_1$-interval $(a_{2i-1}^{n+1}, a_{2i+1}^{n+1})$ and $f(a_{2i}^{n+1})$ as the $\leq^2$-least element of the $\leq_2$-interval $(f(a_{2i-1}^{n+1}), f(a_{2i+1}^{n+1}))$. These selections succeed because both orders are dense. It should be clear that the extended sequence a^{n+1} has the same properties specified for the sequence a^n, so this process can be continued to all values of n. Further, it should be clear that the m-th element of the order $\leq^1$ appears in the domain of f by stage m and the m-th element of the order $\leq_2$ appears in the range of f by stage m, so the definition of f succeeds for all values in the domain of $\leq_1$, defines a function which is onto the domain of $\leq_2$, and is clearly a strictly increasing bijection, so an isomorphism.

Definition: A linear order is said to be *complete* iff every subset of the order which is bounded above has a least upper bound.

Observation: The order on $\mathbb{R}$ is complete.

Theorem: A nonempty separable complete dense linear order with no maximum or minimum is isomorphic to the order on $\mathbb{R}$.

Proof: By the theorem above, the order restricted to the countable dense subset is isomorphic to the usual order on $\mathbb{Q}$, from which it follows easily that the entire order is isomorphic to the usual order on $\mathbb{R}$.

2.18.3 Cardinals and Ordinals; Cardinal Successor; The Hartogs and Sierpinski Theorems

For any cardinal $\kappa < |V|$, there are larger cardinals ($|V|$, for instance). Since the natural order on cardinal numbers is a well-ordering, there is a *smallest* cardinal greater than κ. For finite cardinals n, this next largest cardinal is $n+1$, but of course for infinite κ we have $\kappa = \kappa + 1$: we will see below how the "next" cardinal is obtained in the infinite case.

Definition: If $\kappa \neq |V|$ is a cardinal number, we define $\kappa+$ as the least cardinal in the natural order which is greater than κ.

Now we take an apparent digression into the relationships between cardinal and ordinal numbers. Each ordinal α is naturally associated with a particular cardinal:

Definition: Let α be an ordinal number. We define $\mathtt{card}(\alpha)$ as $|\mathtt{fld}(R)|$ for any $R \in \alpha$ (the choice of R makes no difference).

For each finite cardinal n there is only one ordinal number α such that $\mathtt{card}(\alpha) = n$ (usually written n as well). But for any ordinal α such that $\mathtt{card}(\alpha)$ is infinite, we find that $\mathtt{card}(\alpha + 1) = \mathtt{card}(\alpha) + 1 = \mathtt{card}(\alpha)$: the `card` operation is far from injective. But there is an ordinal naturally associated with each cardinal as well:

Definition: Let κ be a cardinal. We define $\mathtt{init}(\kappa)$ as the smallest ordinal number α such that $\mathtt{card}(\alpha) = \kappa$. There is such an ordinal because any set of size κ can be well-ordered; there is a least such ordinal because the natural order on ordinals is a well-ordering.

It is important to note that the T operations on ordinals and cardinals preserve order, addition, multiplication, and exponentiation. Intuitively, this is all true because $T(\kappa)$ is in some external sense the same cardinal as κ and $T(\alpha)$ is in some external sense the same order type as α. The proofs are straightforward but tedious. One has to take into account the fact that cardinal exponentiation is a partial operation (which reflects the fact that there are more cardinals and ordinals in higher types).

We restate and extend our theorems on the fact that the T operation commutes with operations of arithmetic.

Theorem: Let κ and λ be cardinal numbers. Then $T(\kappa) \leq T(\lambda) \leftrightarrow \kappa \leq \lambda$, $T(\kappa) + T(\lambda) = T(\kappa + \lambda)$, $T(\kappa) \cdot T(\lambda) = T(\kappa \cdot \lambda)$, and $T(\kappa^\lambda) = T(\kappa)^{T(\lambda)}$ if the former exists. $T(\kappa+) = T(\kappa)+$.

Theorem: Let α and β be ordinal numbers. Then $T(\alpha) \leq T(\beta) \leftrightarrow \alpha \leq \beta$, $T(\alpha) + T(\beta) = T(\alpha + \beta)$, $T(\alpha) \cdot T(\beta) = T(\alpha \cdot \beta)$, and $T(\alpha^\beta) = T(\alpha)^{T(\beta)}$ (ordinal exponentiation is total).

We now prove a theorem characterizing the way in which $\kappa+$ is obtained from κ when κ is infinite.

Theorem: Let $\kappa = |A| \neq |V|$ be an infinite cardinal. Let Ω_A be the set of order types of well-orderings of subsets of the set A (clearly this does not depend on the choice of the set A). Then $\kappa+ = \mathtt{card}(\sup(\Omega_A))$.

Proof: Let $\gamma = \mathtt{card}(\sup(\Omega_A))$. Since a well-ordering of a set of size γ must be longer than any well-ordering of a subset of A, $\gamma > \kappa$. Now suppose that $\lambda < \gamma$. It follows that $\mathtt{init}(\lambda) < \mathtt{init}(\gamma) = \sup(\Omega_A)$, so a well-ordering of a set of size λ is of the same length as the well-ordering of some subset of a set of size A, so $\lambda \leq \kappa$. Note that the size of the set of ordinals less than $\sup(\Omega_A)$ is $T^2(\gamma)$, so we could also define γ as $T^{-\mathbf{2}}(|\mathtt{seg}_{\leq_\Omega}(\sup(\Omega_A))|)$.

In the absence of Choice the argument above does not work, but there is still something interesting to say.

Definition: For any cardinal $\kappa = |A| \neq |V|$, define Ω_A as the set of order types of well-orderings of subsets of A and $\aleph(\kappa)$ as $\mathtt{card}(\sup(\Omega_A))$.

Observation: The preceding definition is only of interest in the absence of Choice, as otherwise it coincides with $\kappa+$. Note that $\aleph(\kappa)$ is always a cardinal whose elements are well-orderable. Note that for syntactical reasons this use of $\aleph$ is distinguishable from another use to be introduced shortly.

Theorem (not using Choice; Hartogs): For any cardinal κ, $\aleph(\kappa) \not\leq \kappa$.

Proof: Suppose otherwise. Let $\kappa = |A|$. We then have an injection from a set B of order type $\sup(\Omega_A)$ into A. The range of this injection supports a well-ordering of type $\sup(\Omega_A)$. But the range of this injection is a subset of A, so its order type belongs to Ω_A. This is a contradiction.

Theorem (not using Choice; Sierpinski): $\aleph(\kappa) \leq \exp^3(\kappa)$.

Proof: Since we are working in choice-free mathematics, it is advantageous to represent things in different ways. Any well-ordering is represented effectively by the set of its initial segments. We refer to such a representation of an order as a segment-ordering. A segment-ordinal is an equivalence class of segment-ordinals. Notice that a segment-ordinal is three types higher (not two types higher) than the elements of its field. If $A \in \kappa$, observe that the set of segment-ordinals of well-orderings of subsets of A is of cardinality $T^3(\aleph(\kappa))$. Of course a segment-ordinal is a set of sets of sets of elements of A: the collection of sets of sets of sets of elements of A is of cardinality $T^3(\exp^3(\kappa))$. The desired inequality follows.

A related result is $\aleph(\kappa) \leq \exp^2(\kappa^2)$ This is obtained by noting that the usual ordinals of well-orderings of subsets of A are sets of sets of pairs of elements of κ, so $T^2(\aleph(\kappa)) \leq T^2(\exp^2(\kappa^2))$. This is most useful when we know that $\kappa^2 = \kappa$: this is not a theorem of choice-free mathematics, though it is true if elements of κ are well-orderable or if κ is of the form $\exp^4(\lambda)$ for λ infinite (this last because the construction of the Quine pair can then be mimicked in a set of size κ).

2.18.4 Hierarchies of Cardinals; A Disadvantage of Strong Extensionality

We introduce two notations for cardinal numbers.

Definition: Let $\aleph$ be the natural order on infinite cardinals. We then define $\aleph_\alpha$ for ordinals α using the definition of ordinal indexing of the elements of the field of a well-ordering.

Definition: We define ω_α as $\mathtt{init}(\aleph_\alpha)$.

Definition: Let $\beth$ be the natural order on cardinal numbers restricted to the smallest set of cardinal numbers which contains $\aleph_0$, is closed under the power set operation, and contains suprema of all of its subsets. We then define $\beth_\alpha$ for ordinals α using the definition of ordinal indexing.

We can now pose one of the notable questions of set theory, dating to the beginnings of the subject. The first infinite cardinal is $\aleph_0$. We know by Cantor's Theorem that $|\mathcal{P}(\mathbb{N})| > |\iota``\mathbb{N}| = \aleph_0$. We also note that $|\mathcal{P}(\mathbb{N})| = \beth_1$. We know by definition of cardinal successor that $\aleph_1 = \aleph_0+ > \aleph_0$. We know by the observation following the theorem above that $\aleph_1$ is the number of finite and countable ordinals (which is easily shown to be the same as the number of countable ordinals). The question that arises is the status of

***Cantor's Continuum Hypothesis:** $\beth_1 = \aleph_1$? Are there more subsets of the natural numbers than countable order types?

It is called the Continuum Hypothesis because Cantor also knew (as we will find in the next section) that $\beth_1$ is not only the cardinality of the the set of subsets of the natural numbers but also the cardinality of the set of real numbers, or the number of points on a line (the cardinality of the continuum). For this reason $\beth_1$ is also called c (for "continuum").

A related assertion (which is again a hypothesis not a theorem) is

***Generalized Continuum Hypothesis (GCH):** $\aleph_\alpha = \beth_\alpha$ for all ordinals for which $\aleph_\alpha$ is defined.

A further question is how far the $\aleph_\alpha$'s or $\beth_\alpha$'s continue. These notations are definitely undefined for sufficiently large ordinals α (neither is defined for $\mathtt{ot}(\aleph)$, by a simple consideration of how ordinal indexing is defined). We

cannot prove in this system that $\aleph_\omega$ is defined or even that $\aleph_n$ exists for each natural number n. It is true that $\beth_n$ exists among the cardinals of type n sets for each n, but there is a kind of pun going on here. It is also true that the sequences of $\aleph$'s and $\beth$'s get longer in higher types. Suppose $|V| = \aleph_\alpha$ (there will be such an α). It follows that $T(|V|) = \aleph_{T(\alpha)}$ in the next higher type, so the strictly larger cardinal $|\mathcal{P}(V)| \geq \aleph_{T(\alpha)+1}$, so the sequence is extended in length by at least one. A similar argument for the $\beth_\alpha$'s is slightly more involved.

With strong extensionality there is a much stronger restriction. Suppose that the cardinality of type n is $\aleph_\alpha$. It follows that the largest $\beth_\beta$ which is the cardinality of a type $n+1$ set has $\beta \leq \alpha$. Further, it follows that the largest $\beth_\alpha$ which is the cardinality of a type $n+2$ set has β no greater than $T(\alpha)+2$. Iteration of this observation (and the natural identification of ordinals of different types via the T operation) allow us to say somewhat loosely that there can be no $\beth_\beta$ in any type with $\beta \geq \alpha + \omega$. The reason for this restriction is that there is a definable bound on the size of each type $n+1$ in terms of the size of type n.

This gives a concrete motivation for the form of the axiom of extensionality that we have chosen to use. We do not want the size of mathematical structures that we can consider to be strongly bounded by the size of type 0.

With weak extensionality we can cause much larger $\beth$ numbers to exist because we can assume that each type $n+1$ is much much larger than the power set of type n (a sufficiently large set of urelements is added to support whatever construction we are considering). A strong assumption which suggests itself is that we can iterate the cardinal exponentiation operation on cardinals of sets of type n objects along any well-ordering of type n objects (for each type n). This would give existence of $\beth_{T^{\mathbf{2}}(\alpha)}$ for each ordinal α.

It is useful to note that if we use the convention of type-free cardinal and ordinal variables outlined above, we can treat the exponential operation on cardinals as total. This is achieved in the underlying translation to typed language by providing that we work in a type higher than that of any variable appearing in an exponentiation: the exponential κ^λ is then in effect read as $T(\kappa)^{T(\lambda)}$, which is always defined.

This means that we can in effect say "For every cardinal there is a larger cardinal" and "For every ordinal there is a larger ordinal". $(\forall\kappa.(\exists\lambda.\lambda > \kappa))$ do not make sense under the convention, because we have not bounded the quantifiers. But $(\forall\kappa < \mu.(\exists\lambda < 2^\mu.\kappa < \lambda))$ is true (for any specific μ, with the convention ensuring that we work in a type where 2^μ exists), and expresses

the desired thought.

2.19 Sets of Reals

topological stuff?

2.20 Complex Type Theories

complicated type theories and how they can be represented in TSTU; Curry-Howard isomorphism stuff, perhaps.

2.21 Infinite Indexed Families; König's Theorem

2.22 Partition Relations

We begin this section by stating an obvious

Theorem: If X is an infinite set, A is a finite set, and $f : X \to A$, then there is $a \in A$ such that $f^{-1}\text{“}\{a\}$ is an infinite subset of X.

Proof: The preimages of individual elements of A under f are a disjoint finite family of sets covering X. The sum of their cardinalities is the cardinality of X. If all of them had finite cardinality, this sum (the cardinality of X) would be finite. But the cardinality of X is infinite by assumption.

The first major theorem of this section is a generalization of this.

Definition: If X is a set and κ is a cardinal, we define $[X]^\kappa$ as $\mathcal{P}(X) \cap \kappa$, the set of all subsets of X of size κ.

Definition: If X is a set, n is a natural number, A is a finite set, and $f : [X]^n \to A$, we say that H is a homogeneous set for f iff $H \subseteq X$ and $|f\text{“}[H]^n| = 1$.

Theorem (Ramsey): If X is an infinite set, n is a natural number, A is a finite set, and $f : [X]^n \to A$, there is an infinite homogeneous set H for f.

Proof: For $n = 1$, the result follows immediately from the first theorem of this section.

Assume that the theorem is true for $n = k$ and show that it follows for $n = k + 1$. Let X be an infinite set, A a finite set, and $f : [X]^{k+1} \to A$ a function. Our goal is to show that there is an infinite homogeneous set H for f.

We define a tree T_f. We well-order X and we assume that $u\,T_f\,v$ is defined for all $u, v \leq x$. We define $x\,T_f\,u$ as false for $u < x$. We define $y\,T_f\,x$, for $y < x$, as true iff for all k-element subsets K of $\mathtt{seg}_{T_f}(y)$, $f(K \cup \{y\}) = f(K \cup \{x\})$. $\leq_f$ is a tree because the order on any segment in $\leq_f$ agrees with the underlying well-ordering of X.

We introduce some terminology useful in the context of trees. The *level* of an element x of the field of a tree $\leq_T$ is the order type of $\mathtt{seg}_{\leq_T}(x)$. The *branching* of the tree at an element x of its field is the cardinality of the set of all y such that x is the maximal element of the segment determined by y in the tree. Such elements y are called successors of x in the tree.

In the tree $\leq_f$, the branching will be finite at any element of the field of the tree at finite level. There will be nontrivial branching above an element y just in case there are elements z, w such that for any k-element subset K of $\mathtt{seg}_{\leq_f}(y)$, $f(K \cup \{y\}) = f(K \cup \{z\}) = f(K \cup \{w\})$, but for some $k-1$-element subset A, $f(A \cup \{y, z\}) \neq f(A \cup \{y, w\})$. A possible new branch above y is determined by the assignment of a value under f to each $f(A \cup \{y, z\})$, where z is the next element of the branch. Since there are finitely many subsets of the segment determined by y (since its level is finite) and finitely many values in A, the branching at each element of finite level is finite. One can further prove that if the branching at each element of a finite level is finite, each finite level is a finite set. It follows that some element of each finite level is dominated by infinitely many members of X in the tree order, and further that if some element of finite level is dominated by infinitely many elements of X, it has a successor that is dominated by infinitely many elements of X. From this it follows that we can construct a branch of the tree

with the property that each of its elements of finite level is dominated by infinitely many elements of X (so it has elements of all finite levels and is infinite).

Any branch B in the tree $\leq_f$ has the property that if $b_1 \leq_f b_2 \leq_f \ldots b_k \leq_f b_{k+1} \leq_f c$, that $f(\{b_1, b_2, \ldots, b_k, b_{k+1}\}) = f(\{b_1, b_2, \ldots, b_k, c\})$: the value at a $k+1$-element subset of the branch is not changed if the top element of the set is changed. Thus we can define a new function $f^* : [B]^k \to A$ by $f(\{b_1, b_2, \ldots, b_k\}) = f(\{b_1, b_2, \ldots, b_k, c\})$ for any $c \geq_f b_k$. Now let B be the infinite branch whose existence was shown above. By inductive hypothesis, there is an infinite homogeneous set H for B with respect to f^*, which will also be an infinite homogeneous set for f. This completes the proof.

Ramsey theorem and Erdös-Rado theorem: not part of the main agenda, used for model theory of alternative set theories later.

The Schmerl partition relations, needed for theory of NFUA.

2.23 Large Cardinals

inaccessibles, Mahlos, weakly compact and measurables explained. This is prerequisite knowledge for the model theory of strong extensions of *NFU*; it can also be used to talk about model theory of *ZFC*.

2.24 Pictures of Sets: the Theory of Isomorphism Types of Well Founded Extensional Relations

In this section, we show how the type theory we are working in can naturally motivate a development of the untyped set theory which is more often used, as the theory of a quite natural class of mathematical structures which has its own intrinsic interest.

2.24.1 Coding Sets in Relations

We consider the possibility that a set relation R may be used to represent the membership "relation" $\in$. Toward this end, we introduce some definitions.

Definition: Let R be a relation. We say that an element x of $\mathtt{fld}(R)$ *codes* the set $R^{-1}\text{“}\{x\} = \{y \mid y\,R\,x\}$ relative to R. (if the relation is understood in the context we may just say that the element x codes the given set).

The Definition ensures that a given domain element codes just one subset of the field of the relation, but we would also like it to be the case that a given set is coded by no more than one domain element.

Definition: A relation R is said to be *[weakly] extensional* iff for all x and y in the field of R, if $R^{-1}\text{“}\{x\} = R^{-1}\text{“}\{y\}$ then [either $R^{-1}\text{“}\{x\} = R^{-1}\text{“}\{y\} = \emptyset$] or $x = y$.

A weakly extensional relation leaves open the possibility of coding a theory of sets with distinct urelements, such as are allowed to exist in our type theory: there may be many distinct R-minimal objects if R is weakly extensional, but only one if R is extensional.

Because we are working with set relations, we perforce are at least tempted to use untyped language. For example, we can ask the question whether there is a code for the set $\{x \in \mathtt{fld}(R) \mid \neg x\,R\,x\}$ relative to the relation R. The argument for Russell's paradox shows us that there cannot be such a code (though the set certainly exists). In our type theory we cannot even ask the question which leads to Russell's paradox.

A notion which is difficult (though not entirely impossible) to develop in type theory is the notion of the collection of elements of a set, elements of its elements, elements of elements of its elements, and so forth (a kind of downward closure). In the theory of coded sets this is straightforward.

Definition: Let R be a relation (we do not require it to be extensional). Let x be an element of the field of R. We define the *component* of x determined by R as $R \cap D_x(R)^2$, where $D_x(R)$ is the minimal subset of the field of R which contains x as an element and contains $R^{-1}\text{“}\{y\}$ as a subset for each of its elements y. We denote the component of x determined by R by $C_x(R)$.

Theorem: Let R^* be the minimal reflexive, transitive relation which includes R. Then $C_y(R)$ is $R \cap \{x \mid x\, R^*\, y\}^2$.

Proof: $x \in D_x(R)$ is obvious. Suppose $x \in D_y(R)$ and $y \in D_z(R)$. Any set which contains z as an element and which includes $R^{-1}\text{“}\{u\}$ as a subset for each of its elements u must contain y (by definition of $D_z(R)$ and the fact that $y \in D_z(R)$) and so further must contain x (by definition of $D_y(R)$ and the fact that $x \in D_y(R)$) so we have shown that $x \in D_z(R)$. Thus the relation $x\, S\, y$ defined as $x \in D_y(R)$ is reflexive and transitive, so $x\, R^*\, y$ implies $x \in D_y(R)$. Now observe that $\{y \mid y\, R^*\, x\}$ contains x and includes the preimage under R of any of its elements, so must be included in $D_y(R)$. We now see that the field $D_y(R)$ of the component $C_y(R)$ is precisely $\{x \mid x\, R^* y\}$, from which the result follows.

There is a notion of isomorphism appropriate to weakly extensional relations.

Definition: If R and S are weakly extensional relations, we say that f is a *membership-isomorphism* from R to S if f is a bijection from the field of R to the field of S such that $x\, R\, y \leftrightarrow f(x)\, S\, f(y)$ and in addition if $R^{-1}\text{“}\{x\} = S^{-1}\text{“}(f(x)) = \emptyset$ it also follows that $x = f(x)$.

We impose a further condition on relations which we regard as simulating the membership relation, for which we need to supply a motivation.

Definition: A *[weak] membership diagram* is a well-founded [weakly] extensional relation.

Theorem: If R is well-founded, so is $R^* - [=]$.

Proof: Suppose A is a nonempty subset of $\texttt{fld}(R^* - [=])$ with no $(R^* - [=])$-minimal element. Certainly A is a nonempty subset of $\texttt{fld}(R)$. Let a be an R-minimal element of A. There must be $b \neq a$ such that $b\, R^*\, a$ (since there is no $(R^* - [=])$-minimal element). But from $b\, R^*\, a$, it is easy to deduce $(\exists x.x\, R\, a)$, which is a contradiction.

The effect of the well-foundedness restriction is to ensure that if R and S are membership diagrams and f is a membership-isomorphism from R to S, we can be certain that x with respect to R and $f(x)$ with respect to S always "represent precisely the same set". It is somewhat difficult to say precisely what is meant by this (since we do not yet have an independent understanding of untyped set theory), but a definite result which we can state is that the membership-isomorphism f is *unique*: there can be no other membership-isomorphism from R to S. Suppose there was another such membership-isomorphism g. There would be an R-minimal x in the domain of R such that $f(x) \neq g(x)$. If the R-preimage of x were empty, then so would be the S-preimages of $f(x)$ and $g(x)$, but further we would have $x = f(x) = g(x)$, contradicting the choice of x as a counterexample. If the R-preimage of x were a nonempty set A, then the S-preimage of $f(x)$ would be $f\text{“}A$ and the S-preimage of $g(x)$ would be $g\text{“}A$. But by minimality of x, $f\text{“}A = g\text{“}A$, so by extensionality of S, $f(x) = g(x)$, contradicting the choice of x as a counterexample.

The informal argument that each element of x designates the same set relative to R that is designated by $f(x)$ with respect to S has the same form, but has an essential vagueness dictated by the fact that we are not actually previously acquainted with the domain of sets being designated. If the R-preimage of x is empty, then $x = f(x)$: the two objects represent the same atom. If the R-preimage of x is a nonempty set A, then the S-preimage of $f(x)$ is $f\text{“}A$. x designates (with respect to R) the collection of things designated by elements of A with respect to R. By the minimality hypothesis, the things designated with respect to S by elements of $f\text{“}A$ are the same: so $f(x)$ designates the same collection with respect to S that x designates with respect to R.

Further, if we are working with relations that are extensional rather than weakly extensional, the argument above works with isomorphism in place of membership-isomorphism.

General well-founded relations can be "collapsed" to well-founded [weakly] extensional relations in a suitable sense.

Theorem: Let R be a well-founded relation. Then there is a uniquely determined equivalence relation $\sim$ on $\mathtt{fld}(R)$ with the following property (in which we use the notation $[x]$ for $[x]_\sim$): the relation $R_\sim = \{\langle [x], [y]\rangle \mid x\,R\,y\}$ is [weakly] extensional and for each $[x]$ we have the set of its $R_\sim$-preimages exactly the set of $[y]$ such that $y\,R\,x$.

Proof: Let x be minimal in R such that $C_x(R)$ does not have this property. (Clearly if there is no such x, then the unions of uniquely determined equivalence relations on all $C_x(R)$'s with the indicated property will give such an equivalence relation on R.) Each $C_y(R)$ for $y\,R\,x$ will support such a unique equivalence relation, if it is nonempty. We define the desired equivalence relation on $C_x(R)$, contrary to hypothesis. The top x is equivalent only to itself. All R-preimages of x which have empty R-preimage are either equivalent only to themselves (if we are working with membership-isomorphism) or equivalent to all such preimages (if we are working with isomorphism). Each other element y of $C_x(R)$ has an associated equivalence relation $\sim$ and relation $C_y(R)_\sim$: define $y \sim z$ as holding if and only if $C_y(R)_\sim$ is [membership]-isomorphic to $C_z(R)_\sim$. By hypothesis the restriction of the equivalence relation to each proper component is unique. Extensionality (and the known uniqueness of isomorphisms between well-founded [weakly] extensional relations] leaves us no freedom of choice with respect to defining the equivalence relation between elements of different components. So the equivalence relation obtained is unique.

To see why non-well-founded "membership diagrams" are problematic, consider a diagram containing two elements x and y, each related just to itself. This codes two sets, each of which is its own sole element. Consider another diagram containing two elements u and v, each related just to itself. Either of the two bijections between the fields of these relations is a membership-isomorphism (and indeed an isomorphism) between the relations: there is no way to determine whether x is to be identified with u or with v.

It should be noted that non-well-founded "membership diagrams" are *merely* problematic, not impossible. Interesting untyped theories can be developed in which there are objects which are their own sole elements (and

in which there can be many such objects), and in fact we will have occasion to see this later. Indeed, arbitarily complex failures of well-foundedness of the membership relation are possible and worthy of study.

2.24.2 Passing to Isomorphism Types

The advantage of restricting ourselves to well-founded [weak] membership diagrams is that for any element x of the field of a well-founded membership diagram R, the intended reference of x is in effect fixed by the [membership-]isomorphism type of the component $C_x(R)$. We can then view the [membership-]isomorphism types of components of diagrams as the actual objects under study. When studying weak membership diagrams, there is an element of arbitrariness in the choice of atoms, though it is sometimes useful to have atoms in untyped set theory. The isomorphism types of well-founded extensional relations will be our principal study, and we will see that they correspond precisely to the objects of the usual untyped set theory, though without strong assumptions we will not see the *entire* universe of the usual set theory [in whatever sense this is possible].

Observation: If a [weak] set diagram R is equal to $C_x(R)$ and to $C_y(R)$ where x and y belong to the field of R, then $x = y$. This condition implies $x\,R^*\,y$ and $y\,R^*x$. Since $R^* - [=]$ is well-founded, an $(R^* - [=])$-minimal element of $\{x, y\}$ must be equal to both x and y, so $x = y$.

Definition: A *weak set diagram* is a weak membership diagram which is equal to one of its components (and thus must be nonempty). A *set diagram* is a membership diagram which is either empty or equal to one of its components. A *top* of a [weak] set diagram is either the unique x such that the diagram is its own component determined by x or (in case the diagram is empty) any object whatsoever. A *[weak] set picture* is the [membership-]isomorphism class of a [weak] set diagram [or a double singleton (representing an atom)]. The set of all set pictures is called Z. The set of all weak set pictures whose elements have atoms restricted to a set A is called $Z[A]$ (this last will contain only double singletons of elements of A; of course $Z[V]$ contains all weak set pictures).

Definition: For any [weak] set diagram R with top t, we define an *immediate component* of R as a component $C_x(R)$ such that $x\,R\,t$. Note that the empty set diagram has no immediate components, but may occur as an immediate component of a set diagram if the $x\,R\,t$ happens to have empty R-preimage: the handling of elementless objects in weak set diagrams is seen below. For set pictures ρ and σ, we define $\rho\,E\,\sigma$ as holding iff there are $R \in \rho$ and $S \in \sigma$ such that R is an immediate

component of S. For weak set pictures ρ and σ, we define $\rho\, E\, \sigma$ as holding iff there are $R \in \rho$ and $S \in \sigma$ such that R is an immediate component of S, or ρ is a double singleton $\{\{x\}\}$, and σ has an element S with top t such that $x\, S\, t$ and the S-preimage of x is empty (this handles atoms). It is important to note that no double singleton is a membership-isomorphism class of weak set diagrams, so there is no conflict between the two parts of the definition of E on weak set pictures (the double singleton of the empty set is a set picture, and the sole elementless object in the "set theory" implemented using set diagrams).

Theorem: E is a membership diagram (on Z or $Z[A]$).

Proof: We need to show that E is [weakly] extensional and that E is well-founded. Suppose that ρ and σ are [weak] set pictures and $E^{-1}\text{“}\{\rho\} = E^{-1}\text{“}\{\sigma\}$. This means that each immediate component of any $R \in \rho$ is isomorphic to some immediate component of any $S \in \sigma$ and vice versa. [In the weak case, any preimage of the top of R which has empty R-preimage is identical to some preimage of the top of S which has empty S-preimage, and vice versa]. There is a unique isomorphism from the field of each immediate component of R to a uniquely determined immediate component of S (because no two distinct immediate components can be isomorphic). Any two of these isomorphisms will agree on any common element of their domains. It follows that the union of these isomorphisms, taken together with the pair whose first projection is the top of R and whose second projection is the top of S, yields a [membership-]isomorphism from R to S, so $\rho = \sigma$. [The fact that it is a membership-isomorphism in the weak case follows from the bracketed complete sentence above: elements of $E^{-1}\text{“}\{\rho\}$ and $E^{-1}\text{“}\{\sigma\}$ which are double singletons each correspond to identical elements of the other, and this allows one to define the isomorphism so that it fixes all elements with empty E-preimage]. We have shown that E is [weakly] extensional.

Suppose that A is a nonempty subset of the field of E. Let ρ be an element of A and let $R \in \rho$. [If R is a double singleton, R is E-minimal and we are done.] Define A_R as the intersection of A with the set of isomorphism types of components $C_x(R)$ [and double singletons of R-minimal elements of the field of R]. There will be a minimal x such that the isomorphism type of $C_x(R)$ belongs to A [or there will be a

double singleton which belongs to A]; the isomorphism class of $C_x(R)$ [or the double singleton] will be an E-minimal element of A_R, and so an E-minimal element of A.

Observation: Note that E is two types higher than the [weak] membership diagrams R with which we started. If x in the field of R is at type k, then R itself is at type $k+1$, the [membership-]isomorphism class of R is at type $k+2$, and E is at type $k+3$. We see that E is two types higher than the arbitrary membership diagrams with which we started. E is a kind of universal membership diagram, but this type differential will allow us to completely naturally evade any supposed paradoxical consequences of this universality. The situation here is analogous to that for ordinals: the well-ordering on all ordinals is a kind of universal well-ordering – it contains not a suborder isomorphic to each well-ordering R but a suborder isomorphic to the double singleton image $R^{\iota\mathbf{2}}$ of each well-ordering R. It is also worth noting that strict well-orderings with maxima (and the empty strict well-ordering) are well-founded extensional relations, so there are elements of Z (or $Z[A]$) naturally related to the ordinal numbers (and indeed these correspond precisely to the objects (the *von Neumann ordinals*) which are normally taken to be the ordinal numbers in the usual set theory). One must observe though that a nonzero ordinal α is implemented in untyped set theory by the isomorphism class of the strict well-ordering derived from a well-ordering of order type $\alpha+1$.

There is a type-shifting operation T on [weak] set pictures analogous to the operations on cardinals and ordinals.

Definition: For any [weak] set diagram R, define R^{ι} as usual: this will still be a [weak] set diagram. Let ρ be the [membership-]isomorphism class of R: then $T(\rho)$ is defined as the [membership]-isomorphism class of R^{ι}, and it is straightforward to show that the specific choice of an element R of ρ has no effect on the definition of $T(\rho)$. Notice that in the case of weak set diagrams, atoms are replaced by their singletons as we pass up one type. [Define $T(\{\{x\}\})$ as $\{\{\{x\}\}\}$].

Theorem: For all [weak] set pictures ρ and σ, $\rho\, E\, \sigma \leftrightarrow T(\rho)\, E\, T(\sigma)$.

Proof: This follows directly from the precise parallelism of the structure of $S \in \sigma$ with the structure of $S^\iota \in T(\sigma)$. If $\rho\, E\, \sigma$, any $S \in \sigma$ has an immediate component $R \in \rho$, so belonging to ρ: it is immediate that $S^\iota \in T(\sigma)$ has an immediate component R^ι belonging to $T(\rho)$, so $T(\rho) \in T(\sigma)$. Suppose $T(\rho) \in T(\sigma)$. Then we can choose an element of $T(\sigma)$ of the form S^ι where $S \in \sigma$, which will have an immediate component $R^\iota \in T(\rho)$ (any component of S^ι is obviously a relation singleton image), from which we discover $R \in \rho$, so $\rho \in \sigma$. [If the top of $S \in \sigma$ has an immediate preimage x with empty S-preimage, and $\rho = \{\{x\}\}$, then the top of S^ι has an immediate preimage $\{x\}$, so $\{\{\{x\}\}\} = T(\rho)\, E\, T(\sigma)$ in this case as well; if $T(\rho) \in T(\sigma)$ where $\rho = \{\{x\}\}$, the top of $S^\iota \in T(\sigma)$ has an immediate preimage $\{x\}$ with empty S^ι-preimage [recall that we can without loss of generality choose an element of σ of the form S^ι], we see that the top of $S \in \sigma$ has the preimage x with empty S-preimage, so $\{\{x\}\} = \rho\, E\, \sigma$].

Theorem: For each $\rho \in Z$ $[Z[A]]$ we have $C_\rho(E) \in T^{\mathbf{2}}(\rho)$.

Proof: Let $R \in \rho$. Define ρ_x as the isomorphism type of $C_x(R)$ for $x \in \mathtt{fld}(R)$ [or as $\{\{x\}\}$ if x is R-minimal.] The ρ_x's are exactly the elements of $D_\rho(E)$, $\rho_x\, E\, \rho_y$ iff $x\, R\, y$, but ρ_x is two types higher than x, so we can define a [membership-]isomorphism sending each $\{\{x\}\}$ to ρ_x, witnessing the desired relation between $R^{\iota^{\mathbf{2}}}$ and $C_\rho(E)$.

Theorem (using Choice): Every subset of $T\text{“}Z[A]$ is coded in E. Every subset of $T\text{“}Z$ is coded in E.

Proof: Let B be an arbitrary subset of $T\text{“}Z$ $[T\text{“}Z[A]]$. Each element of B is of the form $T(\rho)$ We transform each $R^\iota \in T(\rho)$ for each $\rho \in B$ to a different R' still belonging to $T(\rho)$: $R' = \{\langle \langle \{x\}, R\rangle, \langle \{y\}, R\rangle\rangle \mid x\, R\, y\}$. The collection of relations R' is pairwise disjoint, so we can take their union and adjoin all pairs $\langle t, T\rangle$ as new elements, where t is the top of one of the R''s and T is a fixed new top element (any pair whose second projection does not belong to B will do). The resulting relation is well-founded and has immediate components of exactly the right isomorphism classes, but it is not extensional. By the theorem proved above on collapsing well-founded relations to well-founded [weakly] extensional relations, we can define an equivalence relation on its field

and replace each element of the field by a representative of its equivalence class taken from a fixed choice set in such a way as to obtain a [weak] set diagram which has immediate components with isomorphism classes which are all and only the elements of B.

Theorem (not using Choice): Every subset of $T^{\mathbf{2}}\text{“}Z[A]$ is coded in E. Every subset of $T^{\mathbf{2}}\text{“}Z$ is coded in E.

Proof: Let B be a subset of $T^{\mathbf{2}}\text{“}Z\ [T^{\mathbf{2}}\text{“}Z[A]]$. Each element $T^{\mathbf{2}}(\rho)$ of B has a *canonical* representative, namely $C_\rho(E)$. These relations all agree on shared members of their domains (since they are all subsets of E). Add a new top element T and add all pairs $\langle \rho, T\rangle$ for $T^{\mathbf{2}}(\rho) \in B$ as elements to their union to obtain a relation with the correct isomorphism classes of immediate components.

Observation: The membership diagram E in higher types faithfully reproduces the membership diagrams in the E relations in lower types. Moreover, the E relation in higher types is *complete* in an obvious sense on its copy of the domains of the E relation of lower types: it codes all subsets of the domains at lower types, whereas a specific E relation cannot code all subsets of *its own* domain. For example, a specific relation E cannot code its own field $Z = \mathtt{fld}(E)$, because it is a well-founded relation (a code v for the entire field of E would satisfy $v\,E\,v$). But $T\text{“}\mathtt{fld}(E)$ is coded (in E of a higher type) from which we can see that more sets are coded in higher types.

2.24.3 The Hierarchy of Ranks of Set Pictures

We introduce the analogue here of the cumulative hierarchy of sets in the usual set theory – without atoms. From this point on we restrict ourselves to membership diagrams, though the results for weak membership diagrams are quite similar.

Definition: For any set $A \subseteq \mathtt{fld}(E)$, we define $P(A)$ as the set of elements of $\mathtt{fld}(E)$ which code subsets of A. We say that the subset A is *complete* if $P(A)$ contains codes for all subsets of A. Notice that $P(A)$ has the same type as A.

Definition: We define the set of *ranks in* E as the intersection of all sets H such that $\emptyset \in H$, $(\forall h \in H.P(h) \in H)$, and $(\forall A \subseteq H.\bigcup A \in H)$.

Theorem: The set of ranks itself contains $\emptyset$, is closed under P and closed under unions of sets of ranks. The ranks in E are well-ordered by inclusion.

Theorem: $\mathtt{fld}(E)$ is a rank.

Definition: Let $\mathbb{E}$ denote the inclusion order on ranks in E. Then $\mathbb{E}_\alpha$ is a general notation for ranks using our convention on ordinal indexing.

Definition: Let γ be the ordinal such that $\mathbb{E}_\gamma$ is the first incomplete rank.

Theorem: $\mathbb{E}_{\omega+n}$ is a complete rank in a high enough type for each familiar natural number n.

Theorem: $|\mathbb{E}_{\omega+\alpha}| = \beth_\alpha$ if $\mathbb{E}_{\omega+\alpha}$ is complete.

The ranks code an iterative process for constructing sets by iterating the "power set" construction which may go through stages indexed by infinite ordinals. This is reminiscent of how the world of our type theory is constructed, except that we lack the ability (or indeed the need) to pass to transfinite levels.[10].

[10]We will explore further the question as to whether type theory suffers from the lack of transfinite levels. But notice that we are able to discuss the transfinite levels of the cumulative hierarchy in type theory here, and the possible presence of urelements means that the hierarchy will not necessarily be truncated at any definite point as it would be in a strongly extensional development of type theory

The set pictures are isomorphism classes supporting a T operation, so we can introduce type free variables ranging over set pictures using the conventions introduced above. Each set picture variable needs to be restricted to some definite type, which can be viewed as restriction of the variable to some set of set picture variables (in higher types) which can in turn be viewed as restriction of the variable to the preimage under E of some set picture (if we go up one more type so that all elements of the original type are images under T so we have completeness). Just as we represented the bounding of ordinal variables in types as bounding in the segment determined by an ordinal variable, we can represent the bounding of set picture variables in types or sets within types as a bounding in the preimage of a set picture under E.

The self-contained theory of set pictures thus obtained is an untyped set theory with E as its membership.

We outline the proofs of some important theorems of this untyped theory.

Theorem: For every set picture σ and every formula ϕ, there is a set picture τ such that $(\forall \rho\, E\, \tau . \rho\, E\, \sigma)$ and $(\forall \rho \in \sigma . \rho \in \tau \leftrightarrow \phi)$.

Proof: Our conventions ensure that we work in a type where $\sigma = T(\sigma')$ for some σ', and the result then follows from theorems given above: the image under T of any set of set pictures is coded.

Theorem: For every set picture σ, the set of all codes of subsets of the preimage of σ under E is coded.

Proof: Just as with the result that cardinal exponentiation is total in the untyped theory of cardinals, this is achieved by clever definition of our conventions. We stipulate that if any set picture σ is mentioned, we work in a type high enough that $\sigma = T(\sigma')$ for some σ'. This ensures that any subcollection of the preimage of σ under E is coded (the burden of the previous theorem) and is further itself also an image under T, so the collection of all these subsets is also coded (though it is not necessarily an image under T). Note that if we further mention this set (for a specific σ) we bump ourselves into a yet higher type (so we can iterate this "power set" operation any concrete finite number of times).

2.25 Category theory

We give a brief introduction to category theory as carried out in type theory.

Definition: A category is a tuple $\langle O, M, d, r, \circ, \mathtt{id}\rangle$, where O is the set of *objects* of the category and M is the set of *morphisms* of the category, $d : M \to O$ is the *domain* function of the category, $r : O \to M$ is the *codomain* function of the category, and $\circ \subseteq (M \times M) \times M$ is the *composition* operation of the category (which is partial, so we do not write $\circ \subseteq M \times M \to M$). The function $\mathtt{id} : O \to M$ is the identity morphism constructor.

Certain conditions must be satisfied which are necessary and sufficient for this to be a category. For each pair of morphisms f, g there is at most one h such that $\langle\langle f, g\rangle, h\rangle \in \circ$, and we write $h = f \circ g$ if it exists. The condition for $g \circ f$ to exist is exactly $r(f) = d(g)$, and $d(g \circ f) = d(f)$, $r(g \circ f) = r(g)$, for any morphisms f, g for which $g \circ f$ is defined. $f \circ (g \circ h) = (f \circ g) \circ h$ whenever the compositions involved exist. $\mathtt{id}(r(f)) \circ f = f \circ \mathtt{id}(d(f)) = f$ for all morphisms f.

We could economize on components in our tuple if we identified each object A with $\mathtt{id}(A)$; d and r would then be functions $M \to M$ and $\mathtt{id}$ and O would not be needed as components. However, in natural examples we do not tend to identify objects with their identity morphisms.

A specific concrete example of a category is the category of sets and functions in any particular type. The objects of this category are all the sets of a type n. The morphisms f of the category with $d(f) = A$ and $r(f) = B$ are just exactly the functions $f : A \to B$. If A is an object, that is a set, $\mathtt{id}(A)$ is the restriction of the identity function to A. The composition operation is the usual operation of composition of functions, restricted appropriately.

The alert reader may notice that we are lying (we are doing so quite deliberately, but will promptly atone). The difficulty is that for each function f we do not have a unique $r(f)$ such that $f : \mathtt{dom}(f) \to r(f)$. We fix this by clarifying that our "functions" are actually pairs $\langle f, B\rangle$ where f is a function in the usual sense and $\mathtt{rng}(f) \subseteq B$, and defining $\mathtt{id}(A)$ as $\langle \mathtt{id}\lceil A, A\rangle$, the identity function restricted to A, paired with A, and $\langle g, C\rangle \circ \langle f, B\rangle$ as $\langle g \circ f, C\rangle$, where the composition on the left is the usual composition of functions. Of course $d(\langle f, b\rangle) = \mathtt{dom}(f)$ and $r(\langle f, B\rangle) = B$. We will call these objects "typed functions" and we may now and then confuse them

with their first projections when their intended codomain is evident from context.

We could also have defined $r(B)$ as $\mathtt{rng}(B)$, but this would have given a different category. Notice that with this definition the set of morphisms from a set A to a set B would be the set of functions from A *onto* B, so we might call this the category of surjective functions (it contains exactly the same functions, but organized differently).

For any objects A, B, we define $\mathtt{hom}(A, B)$ as the set of morphisms f with $d(f) = A$, $r(f) = B$. We call these sets homsets.

Just for fun, I provide an alternative representation of category theory which might indicate that it can be freed from dependency on notions of function per se.

A *multigraph* is a triple $\langle V, E, g\rangle$ where V is the set of vertices of the graph, E is the set of edges, and $g : E \to V \times V$ tells us where each edge starts and ends: if $g(e) = \langle a, b\rangle$, then e is an edge from a to b. A *path* is a multigraph is a finite sequence p of odd length in which p_{2n} is always a vertex and p_{2n+1} is an edge and satisfies $g(p_{2n+1}) = \langle p_{2n}, p_{2n+2}\rangle$. A path p with domain $[0, n]$ is said to be a path from p_0 to p_n. The concatenation $p \oplus q$ of p and q, where p has domain $[0, n]$, q has domain $[0, m]$, and $p_n = q_0$ is defined thus, with domain $[0, m + n]$: $(p \oplus q)_i$ is either p_i or q_{i-n}, as appropriate. A category is then determined by an equivalence relation on paths in a multigraph, with the properties that if $p \sim q$ and p is a path from a to b, q is also a path from a to b, and which respects concatenation: if $p \sim r$ and $q \sim s$, then $p \oplus q \sim r \oplus s$. If the additional condition is imposed that each equivalence class contains exactly one path with domain $[0, 0]$ or $[0, 2]$ (a path determined by a single edge or vertex), there is a precise correspondence between multigraphs and categories (notice that such multigraphs must contain edges from every a to every b for which there is a path from a to b, where $a \neq b$; it may or may not contain loops at each vertex, and loops will not be equivalent to paths with domain $[0, 0]$); in other cases, non-isomorphic multigraphs may generate isomorphic associated categories, basically by adding lots of virtual edges to the underlying multigraph and collapsing some actual edges together. Notice that in this formulation the morphisms of the category, being equivalence classes of paths, are at a higher type than the objects of the path, which are the vertices. In the restricted formulation with a single path in each equivalence class with domain $[0, 0]$ or $[0, 2]$, we can instead use the single edge in the range of that path as the morphism associated with the equivalence class (or the single vertex if there

is no edge), and that is at the same type as the vertices/objects. In the more general case, we note that paths can be represented at the same type as the edges and vertices in them, by building these finite structures using pairing instead of membership, for example, and then representative paths can be chosen from each equivalence class to serve as morphisms.

The objects of a category are often (but not always) structured sets of some sort and the morphisms are often (but not always) functions which preserve this structure. For example, there is a category of all groups, in which the morphisms are homomorphisms, and a category of topological spaces, in which the morphisms are homeomorphisms. A category which is not exactly of this kind is the category whose objects are topological spaces and whose morphisms are equivalence classes of continuous functions under homotopy.

Reflexively and perhaps worryingly, we can define a category of categories. If C and D are categories (with components as above which we will subscript with their names) a functor from C to D is a function F sending objects of C to objects of D and satisfying $r_D(F(A) = F(r_C(A)), d_D(F(A) = F(d_C(A)), F(f \circ_C g) = F(f) \circ_D F(g), F(\mathtt{id}_C(A)) = \mathtt{id}_D(F(A))$. A functor preserves category theoretic structure. Notice that the image of the functor may not be all of D and the map F may not be an injection. Now, there is a category of categories whose objects are all the categories and whose morphisms are the functors between categories.

We haven't told the whole story here! The alert reader should notice that the category of all type n categories must actually be of type $n + 1$, since it is a tuple one of whose components is the set of type n categories. The concrete example given above, the category of type n sets and functions, is itself a type $n + 1$ category.

A category considered as such (without information about the specific natures of its objects and morphisms) is a sort of infinite diagram with dots (objects) connected by directed arrows (morphisms) and a notion of composition which ensures that any path made up of directed arrows can be identified with a single directed arrow. There is nothing more to it. Properties of categories which are commonly articulated are often of a character which ensures that a category really does have the structure of some kind of system of sets and functions; at any rate, this is the character of the properties we will introduce.

For example, in the category of sets and typed functions, each singleton set $A = \{a\}$ has the property that there is exactly one arrow (we may use the

word "arrow" to mean "morphism") from any object B to A (the constant function with the appropriate value a on B). An object with this property in a general category is called a *terminal object* for that category. The empty set $\emptyset$ has the property that there is exactly one arrow from $\emptyset$ to A for any set A: an object with this property in a general category is called an *initial object*. The category of sets and typed functions has just one initial object and many terminal objects, but there is a sort of uniqueness for terminal objects: for any two terminal objects A and B, there is exactly one arrow from A to B and exactly one arrow from B to A, and their composition must be the unique arrow from B to B, which is $\mathtt{id}(B)$. If $f \circ g = \mathtt{id}(A)$, we say that f and g are inverses, and that f and g are *isomorphisms*. All terminal objects in any category are isomorphic to one another. Similarly, all initial objects in any category are isomorphic to one another. When there is an isomorphism from A to B, composition with the isomorphism gives an exact correlation of structure between homsets involving A and corresponding homsets involving B.

We describe a situation under which we say that a category "has products". For any objects A and B, if we believe that $C = A \times B$, by analogy we expect to have morphisms $\pi_1 : C \to A$ and $\pi_2 : C \to B$ such that for any object D and morphisms $f : D \to A$ and $D \to B$, there is a unique morphism $f \times g : D \to C$ such that $\pi_1 \circ (f \times g) = f$ and $\pi_2 \circ (f \times g) = g$. Notice that the existence of a product of A and B (or of products of any two objects in a category) is not just the assertion of the existence of an object $C = A \times B$ but of the existence of C equipped with "projection maps" π_1 and π_2. It is straightforward to see that any two products of A and B are isomorphic (though there may be more than one of them!) The idea is that $f \times g$ is the function sending $\langle x, y\rangle \in A \times B$ to $\langle f(x), g(y)\rangle$; of course this is uniquely determined and has the indicated property, but in a general category we do not know that the objects are functions in the usual sense. The cute thing about this definition is that it not only singles out the cartesian product of sets (up to a bijection) in the category of sets, but it also picks out a correct notion of product of objects and product of functions in other categories, as for example groups or topological spaces.

Any category C can be transformed into a "converse" category $C^{\mathtt{op}}$, called the *opposite* or *dual* category of C, by reversing the direction of all arrows, and similarly it is often useful to reverse arrows in a property. Suppose that for objects A and B we have an object C such that we have morphisms $\sigma_1 : A \to C$ and $\sigma_2 : B \to C$ and for any object D and arrows $f : A \to D$

and $g : B \to D$, we have a uniquely determined arrow $f + g$ such that $(f + g) \circ \sigma_1 = f$ and $(f + g) \circ \sigma_2 = g$. Curiously, we have described a construction in the standard theory of sets and functions: for any sets A, B, the disjoint union $(A \times \{0\}) \cup (B \times \{1\})$ equipped with the maps $\sigma_1 = (\lambda a \in A : \langle a, 0\rangle)$ and $\sigma_2 = (\lambda b \in B : \langle b, 1\rangle)$ has these properties. It may not seem obvious that cartesian product and disjoint union are "dual" operations, but from the category theoretic standpoint that is how things look.

We continue with a development of "function spaces" internally to a category. We want to say what it means for there to be a category B^A which is in effect the set of functions from A to B. The alert reader will see that we are doing violence to our scheme of types as we work, but we will fix everything up before the end.

The idea for representing B^A using category theory concepts comes from the notion of "currying", popular in computer science for converting functions with two arguments to functions with one argument: where we have a function $f(x, y)$, define a related function $\hat{f}$ such that $\hat{f}(x)(y) = f(x, y)$. We note that to keep types straight we actually need $\hat{f}(\{x\})(y) = f(x, y)$. Doing violence to types which we will duly fix, we want a map $\mathtt{ev}$ which sends a pair $\langle f, a\rangle$ in $B^A \times A$ to $f(a)$ in B. The type fix is the following: we in fact consider ι"A and ι"B and the set B^A and provide the map $\mathtt{ev}$ such that $\mathtt{ev}(\langle f, \{a\}\rangle) = \{f(b)\}$.

We now express the defining property of the exponential C^B and its associated $\mathtt{ev} : C^B \times B \to C$ arrow: For any object A and any arrow $f : A \times B \to C$, there is a unique arrow $\hat{f} : A \to C^B$ such that $\mathtt{ev} \circ (\hat{f} \times 1_B) = f$, where 1_B is the map from B to a convenient terminal object. In our category of sets and typed functions, exponentials C^B exist for any sets of singletons C, B, with C^B implemented as the set of functions from $\bigcup B$ to $\bigcup C$ and the map $\mathtt{ev}$ sending each pair $\langle f, \{b\}\rangle$ to $\{f(b)\}$. The map $\hat{f}$ will be the unique map satisfying $\hat{f}(a)(b) \in f(a, \{b\})$. This is easily implemented for all sets B, C the same size as a set of singletons, as well.

A cartesian closed category is one in which there is a terminal object and there are products and exponentials for every pair of objects. This cannot be the case in the full category of typed sets and functions, as each arrow $f : B \to C$ would of course correspond to a unique arrow $f^* : 1 \times B \to C$ and so would be required to correspond to a unique arrow $\hat{f^*} : 1 \to C^B$: now observe that if $B = C = V$, there are more functions from V to V than there can be functions from a terminal object (singleton set) to any object (there

are more functions from V to V than there are elements of V, and so of any object). This has nothing to do with the specific implementation of cartesian closedness above (which of course does not work for the whole category of typed sets and functions): it is an argument that no implementation can work.

Say that a set A is small if $\mathtt{exp}^n(|A|)$ exists for each n: the category of small sets and typed functions between small sets is cartesian closed. This is a hint as to what the world of sets of untyped set theory has to look like: sets must be restricted to be small in a suitable sense, as we will see in the next chapter.

We define a notion of *small category* motivated by cardinality features of the category of small sets and functions just discussed. Notice that the homsets of the category of small sets and typed functions between them are small, but moreover one has to pay attention to the type relative to which they are small. The category of small sets and functions of type $\mathbf{k+1}$ has homsets at type $\mathbf{k+2}$ whose cardinalities (in type $\mathbf{k+3}$) are images under T of small cardinals of type $\mathbf{k+2}$ (cardinalities of type $\mathbf{k+1}$ sets). So we define a small category as a category whose homsets have cardinalities which are images under T of small cardinals. The appearance of the T operator forces the correct typing. Notice that there is no assertion that the set of objects is bounded in size: it is the collections of arrows between any given pair of objects which are being bounded in size.

An important idea in category theory is that of a *natural transformation*. Given two functors S, T, both from a category C to a category B, a natural transformation from S to T is a function τ from objects of C to morphisms of B such that $\tau(c)$ is a morphism from $S(c)$ to $T(c)$ and for any morphism f in C with $d(f) = c$, $r(f) = c'$ we have $T(f) \circ \tau(c) = \tau(c') \circ S(f)$.

Chapter 3

Untyped theory of sets

In this chapter we introduce the usual untyped set theories (Zermelo set theory and the stronger *ZFC*, as well as some intermediate systems) and relate them to type theory. We will present (at the end) the view that untyped set theory can be interpreted as the theory of set pictures (isomorphism types of certain well-founded extensional relations), which should already be suggested by the treatment at the end of the previous chapter.

Further, we strongly criticize the idea that the axioms of Zermelo set theory are somehow essentially *ad hoc*, as is often suggested (this is stated with great confidence so often as to be cliché). There are some odd features of the earliest form of the axioms, which reflect the fact that they appear early in the process of understanding what can be done with set theory, but Zermelo set theory is very close to being exactly the abstract theory of set pictures, and this is not *ad hoc*. I do think that something is missing from the formulation of Zermelo set theory as an independent theory: adding either the Axiom of Rank or the combination of Foundation and the Mostowski Collapsing Lemma gives a theory with the same mathematical strength and much more satisfactory technical features.

In untyped set theory there is only one kind of object – sets. There may also be atoms if extensionality is weakened to allow them but they will not be an essentially different sort (type) of object. Though this may seem to be quite a different kind of theory, we will see that the usual untyped set theory is not so distantly related to the typed theory of sets we have developed as you might think.

Subsections of this section which depend strongly on the presentation of type theory in the previous section are marked with †, as are local remarks

with such dependencies; subsections which are part of a mostly self-contained treatment of untyped set theory are unmarked.

3.1 The original system of Zermelo

The first modern axiomatic system of set theory was proposed by Zermelo in 1908. It is even older than the first publication of the famous *Principia Mathematica* of Russell and Whitehead, though not as old as Russell's first proposal of the theory of types.

The axioms differ somewhat from those in modern treatments. In this theory, we have primitive predicates of membership and equality, and all objects are of the same sort (there are no type restrictions in our language).

Axiom of Extensionality:

$$(\forall xy.(\exists z.z \in x) \wedge (\forall w.w \in x \leftrightarrow w \in y) \rightarrow x = y)$$

In the statement of this axiom, we follow Zermelo's apparent original intention of allowing atoms. We will usually assume strong extensionality, however. An informal way of putting this axiom is "Non-sets have no elements, and sets with the same elements are equal".

Elementary Sets: There is an object $\emptyset$ such that $(\forall x.x \notin \emptyset)$, called *the empty set*. For any objects x and y, there is a unique object $\{x\}$ such that $(\forall z.z \in \{x\} \leftrightarrow z = x)$ and a unique object $\{x, y\}$ such that $(\forall z.z \in \{x, y\} \leftrightarrow z = x \vee z \in y)$.[1]

Definition: For each formula ϕ, define $\{x \mid \phi\}$ as $\emptyset$ if $(\forall x.\neg\phi)$, and otherwise as the object A such that $(\forall x : x \in A \leftrightarrow \phi)$, if there is such an object. Otherwise we say that $\{x \mid \phi\}$ does not exist[2]. Notice that $\emptyset = \{z \mid z \neq z\}$, $\{x\} = \{z \mid z = x\}$ and $\{x, y\} = \{z \mid z = x \vee z = y\}$.[3]

Define $\texttt{set}(x)$ as $x = \emptyset \vee (\exists y.y \in x)$. It would be equivalent but a little more mysterious to define $\texttt{set}(x)$ as $x = \{y \mid y \in x\}$.

Axiom (scheme) of Separation: For any formula $\phi[x]$ and set A, the set $\{x \mid x \in A \wedge \phi\}$, which we abbreviate $\{x \in A \mid \phi[x]\}$, exists. This can

[1]Notice that the assertion that for any x, y, $\{x, y\}$ exists, implies that all objects are of the same type in the parlance of section 2.1.1.

[2]This can be modified using the device of "proper classes" introduced later.

[3]If we include the Hilbert symbol in our logic, and stipulate that the default object is $\emptyset$, then $\{x \mid \phi\}$ can be defined as $(\epsilon x : (\forall y : y \in x \leftrightarrow \phi)$, if one is willing to live with the odd result that symbols for sets that cannot exist, such as $\{x \mid x \notin x\}$, actually represent $\emptyset$.

be written more rigorously, "For each formula $\phi[x]$,

$$(\forall A.(\exists S.(\forall x.x \in S \leftrightarrow x \in A \wedge \phi[x]))),$$

where S is a variable not appearing in $\phi[x]$".[4]

More complex set builder notation may be used. $\{t[x_1, \ldots, x_n](\in A) \mid \phi\}$ is defined as $\{u(\in A) \mid (\exists x_1, \ldots, x_n.u = t[x_1, \ldots, x_n])\}$, where u is a fresh variable not appearing in ϕ or $t[x_1, \ldots, x_n]$, the latter being shorthand for an arbitrary complex notation containing the x_i's. The notation A for a bounding set (if present) should not depend on x.

Digression: comments on Separation and an alternative: In Zermelo's original formulation, he simply said that a subcollection of a set defined by an arbitrary property was a set. He commented on the usual formulation given above that it was not strong enough to realize his conception: the ways in which subsets can be defined by Separation are constrained by limitations of our language. It turns out that Zermelo set theory with the form of Separation given here (and so of course with stronger forms perhaps closer to his conception) is strictly more powerful than the type theory of chapter 2. On the other hand, a restriction of our language motivated by the form of the axiom of Separation will give a theory precisely equivalent in power to our type theory of chapter 2:

*__Bounded Separation:__ For any set A formula $\phi[x]$ in which each quantifier is bounded, the set $\{x \mid x \in A \wedge \phi\}$, which we abbreviate $\{x \in A \mid \phi[x]\}$, exists. The precise meaning of "each quantifier is bounded" is that each quantifier is of the form $(\forall x \in t. \ldots)$ or $(\forall x \in t. \ldots)$ where t is an expression in which the variable x does not occur.

The motivation here is to apply the same restriction of the range of the main bound variable to a set which appears in the axiom of separation to all occurrences of bound variables in instances of separation.

Axiom of Power Set: For any set A, the set $\{B \mid B \subseteq A\}$ exists. The definition of $A \subseteq B$ is the usual one:

$$A \subseteq B \equiv_{\mathtt{def}} \mathtt{set}(A) \wedge \mathtt{set}(B) \wedge (\forall x.x \in A \rightarrow x \in B).$$

[4]This is technically an "axiom scheme" rather than a single axiom: there is a distinct axiom for each formula $\phi[x]$.

Axiom of Union: For any set A, the set $\bigcup A = \{x \mid (\exists y \in A.x \in y)\}$ exists.

Definition: A set I such that $\emptyset \in I$ and $(\forall x.x \in I \rightarrow \{x\} \in I)$ is said to be Zermelo-inductive.

Axiom of Infinity: There is a set $\mathcal{Z}$ such that $x \in \mathcal{Z}$ iff x belongs to every Zermelo-inductive set. This set is known as the set of *Zermelo natural numbers*: Zermelo used $\emptyset$ to represent 0, $\{0\}$ to represent 1, $\{1\}$ to represent 2, and generally $\{n\}$ to represent $n + 1$.

Definition: We define $A \cap B$ as $\{x \in A \mid x \in B\}$ (and $A - B$ as $\{x \in A \mid x \notin B\}$). Sets A and B are said to be disjoint iff $A \cap B = \emptyset$. A collection $\mathcal{A}$ is said to be *pairwise disjoint* iff $(\forall AB : A \in \mathcal{A} \wedge B \in \mathcal{A} \wedge A \neq B \rightarrow A \cap B = \emptyset)$. A set C is a *choice set* for a pairwise disjoint collection $\mathcal{A}$ iff $(\forall A \in \mathcal{A}.(\exists x.A \cap C = \{x\}))$, i.e., each element of $\mathcal{A}$ shares exactly one element with C.

Axiom of Choice: Any pairwise disjoint collection of nonempty sets has a choice set.

We give some discussion of the axioms. Items in this discussion may presuppose knowledge of our previous discussion of untyped set theory, though general mathematical knowledge may substitute for this.

1. We will usually assume strong extensionality (objects with the same elements are equal), as is now usual, but here we preserve Zermelo's original intention of allowing atoms.

2. The axiom of elementary sets is more complicated than is necessary. The separate provision of the singleton set is not made in the modern treatment, as $\{x\} = \{x, x\}$ exists if we merely assert the existence of unordered pairs, and Separation and Infinity together imply the existence of the empty set ($\emptyset = \{x \in \mathcal{Z} \mid x \neq x\}$) or of at least one empty object if strong extensionality is not assumed.

3. Zermelo did not know that the ordered pair could be defined by $\langle x, y\rangle = \{\{x\}, \{x, y\}\}$, but note that the existence of the ordered pair (now in the Kuratowski form) is provided by the axiom of elementary sets.

4. The axiom of separation does not appear to imply any paradoxes. We attempt the Russell argument: define $R_A = \{x \in A \mid x \notin x\}$. Observe that $R_A \in R_A \leftrightarrow R_A \in A \wedge R_A \notin R_A$. This would only lead to contradiction if $R_A \in A$, so we conclude $R_A \notin A$, whence we conclude that there is no universal set (for every set A we have specified a set R_A which cannot belong to it).

5. The axiom of power set and the axiom of union define familiar constructions. Note that $x \cup y$ can be defined as $\bigcup\{x, y\}$. $x \cap y = \{z \in x \mid z \in y\}$ and $x - y = \{z \in x \mid z \notin y\}$ are provided by Separation alone. Complements do not exist for any set. The cartesian product $A \times B$ is definable as $\{c \in \mathcal{P}^{\mathbf{2}}(A \cup B) \mid (\exists ab.a \in A \wedge b \in B \wedge c = \langle a, b\rangle)\}$.

6. In a modern treatment, the von Neumann successor x^+ is defined as $x \cup \{x\}$, and the axiom of infinity asserts that there is a minimal set which contains the empty set and is closed under the von Neumann successor operation. It is interesting to observe that neither form of the axiom of infinity implies the other in the presence of the other Zermelo axioms (though they are equivalent in the presence of the axiom of replacement introduced below).

7. It is remarkable that in spite of the fact that Zermelo did not know how to code the general theory of relations and functions into set theory (lacking an ordered pair definition) he was able to prove the Well-Ordering Theorem from the Axiom of Choice in his 1908 paper. Some day I have to look at how he did it!

8. The axioms of Foundation and Replacement which complete the modern set theory *ZFC* were later developments.

9. We describe a minimal model of Zermelo set theory. The domain of this model is the union of the sets $\mathcal{P}^{\mathbf{i}}(\mathbb{N})$ (for purposes of this paragraph we take $\mathbb{N}$ to be the set $\mathcal{Z}$ of Zermelo natural numbers). It is important to note that the Zermelo axioms give us no warrant for believing that this sequence of sets makes up a set. Extensionality certainly holds in this structure (in its strong version). The empty set belongs to $\mathbb{N}$, so is certainly found in this structure. It is useful at this point to note that $\mathbb{N} \subseteq \mathcal{P}(\mathbb{N})$ (each Zermelo natural number is a set of Zermelo natural numbers, 0 being the empty set and $n + 1$ being $\{n\}$); since $A \subseteq B$

obviously implies $\mathcal{P}(A) \subseteq \mathcal{P}(B)$, we have (by repeated application) $\mathcal{P}^{\mathbf{i}}(\mathbb{N}) \subseteq \mathcal{P}^{i+1}(\mathbb{N})$ and so $\mathcal{P}^{\mathbf{i}}(\mathbb{N}) \subseteq \mathcal{P}^{j}(\mathbb{N})$ if $i \leq j$. The iterated power sets of the set of natural numbers whose union is our structure are nested. For any x and y in the structure, there are m and n such that $x \in \mathcal{P}^{\mathbf{m}}(\mathbb{N})$ and $y \in \mathcal{P}^{\mathbf{n}}(\mathbb{N})$: both x and y belong to $\mathcal{P}^{\mathbf{m+n}}(\mathbb{N})$, and so $\{x, y\} \in \mathcal{P}^{\mathbf{m+n+1}}(\mathbb{N})$: the structure satisfies the axiom of elementary sets. If $A \in \mathcal{P}^{\mathbf{i}}(\mathbb{N})$, then $\mathcal{P}(A) \in \mathcal{P}^{\mathbf{i+1}}(\mathbb{N})$. If $A \in \mathcal{P}^{\mathbf{i}}(\mathbb{N})$ (for $i > 0$), then $\bigcup A \in \mathcal{P}^{\mathbf{i-1}}(\mathbb{N})$: the restriction to positive i is no real restriction because $\mathbb{N} \subseteq \mathcal{P}(\mathbb{N})$. Infinity obviously holds since $\mathbb{N}$ belongs to the structure. If it is supposed that Choice holds in the whole universe it certainly holds in this structure, as a choice set for a partition in $\mathcal{P}^{\mathbf{i+1}}(\mathbb{N})$ will belong to $\mathcal{P}^{\mathbf{i}}(\mathbb{N})$

† Notice the similarity between the role of iterated power sets of the natural numbers in our description of this structure and types in the theory of the previous chapter. The only difference is that the analogues of types here are cumulative.

3.1.1 Exercises

1. We define x^+ as $x \cup \{x\}$. We use the modern form of the Axiom of Infinity: there is a set which contains $\emptyset$ and is closed under $x \mapsto x^+$. We implement 0 as $\emptyset$, and if the natural number n is implemented as the set x, $n+1$ is implemented as x^+.

   We define $\mathbb{N}$ as the intersection of all sets which contain 0 and are closed under successor. Explain how we can show that this set exists using the axioms of infinity and separation.

   Show that the axioms of Peano arithmetic are satisfied in this implementation of $\mathbb{N}$. Proofs of axioms 1,2,3,5 should be very straightforward.

   Axiom 4 requires you to show that $x \cup \{x\} = y \cup \{y\}$ implies $x = y$ for all $x, y \in \mathbb{N}$. Show this using the axioms of Zermelo set theory (*without* Foundation).

   Hints: how do you prove *anything* about natural numbers? You can begin as an exercise by proving that for no natural number n is $n \in n$ true, by induction of course. This is similar to the fact about natural numbers you need to prove to establish Axiom 4. I will give more explicit hints if you visit me with work in progress.

2. Write a proof in Zermelo set theory with the modern form of the Axiom of Infinity (and without Foundation) that no natural number is an element of itself. This will of course be an induction proof using the definitions $0 = \emptyset; n+1 = n^+ = n \cup \{n\}$. Intense attention to "obvious" detail is needed at this level. Hint: it will be useful (and easy) to prove first (by induction of course) that all natural numbers are transitive: a set A is said to be transitive iff all elements of elements of A are also elements of A.

   Even more of a hint: the induction step looks like this. Suppose $n \notin n$. Our goal is to show $n+1 = n \cup \{n\}$ is not an element of itself. Suppose otherwise for the sake of a contradiction. We suppose that is that $n+1 \in n+1 = n \cup \{n\}$. So either $n+1 \in n$ (something bad happens...) or $n+1 = n$ (something bad happens...).

3.2 Basic set constructions in Zermelo set theory

In this section we develop basic mathematical constructions in Zermelo set theory.

We begin with a very basic

Theorem: $(\forall A.(\exists x.x \notin A))$

Proof: This theorem follows from Separation alone. Consider

$$R_A = \{x \in A \mid x \notin x\}.$$

Suppose $R_A \in A$. It follows that $R_A \in R_A \leftrightarrow R_A \notin R_A$.

It will be seen to follow from the Axiom of Foundation usually added to Zermelo set theory that $x \notin x$ for any x, it further follows from Zermelo set theory with the Axiom of Foundation that $R_A = A$ for all A.

It is worth noting that this is important in applying the Axiom of Infinity. It might seem to be a hazard that there might be no Zermelo-inductive set. But if there were no Zermelo-inductive set, then the set $\mathcal{Z}$ of all sets which belong to every Zermelo-inductive set would contain *everything*, because each object belongs to all elements of the empty set. But there is no set which contains every object. So there is a Zermelo-inductive set, and then one argues in a standard way that $\mathcal{Z}$ itself is the smallest Zermelo-inductive set. Notice though that it is also possible to prove quite directly that $\mathcal{Z}$ is Zermelo-inductive.

It is a fundamental characteristic of Zermelo set theory (and of all stronger theories) that there are no very big sets (such as the universe V). Many mistake this for a fundamental characteristic of set theory.

We want to implement relations and functions. Here it is very convenient to work with the Kuratowski pair.

Definition: $\langle x, y\rangle = \{\{x\}, \{x, y\}\}$

Theorem: For any sets x and y, $\langle x, y\rangle$ is a set. If $\langle x, y\rangle = \langle z, w\rangle$ then $x = z$ and $y = w$.

Definition: $\pi_1(\langle x, y\rangle)$ is defined as the unique u belonging to all elements of $\langle x, y\rangle$: note $\pi_1(\langle x, y\rangle) = x$. $\pi_2(\langle x, y\rangle)$ is defined as the unique u belonging to exactly one element of $\langle x, y\rangle$: note $\pi_2(\langle x, y\rangle) = y$. For a different definition of the ordered pair, these projection operators will be defined differently, in order to satisfy the same equations.

This theorem is not enough by itself to ensure that we can use the Kuratowski pair to get an adequate theory of relations.

Definition: $A \cup B = \bigcup\{A, B\}$

Theorem: $A \cup B$ exists for any sets A and B (this is clear from the form of the definition). $A \cup B = \{x \mid x \in A \vee x \in B\}$. Notice that the latter definition, which we used as the primary definition in type theory, is *not* guaranteed to define a set by Separation.

Definition: $A \cap B = \{x \in A \mid x \in B\}$; $A - B = \{x \in A \mid x \notin B\}$. If A is a nonempty set and $B \in A$, $\bigcap A = \{x \in B \mid (\forall a \in A.x \in A)\}$.

Definition: For any natural number $n > 2$, we define $\{x_1, x_2, \ldots, x_n\}$ as $\{x_1\} \cup \{x_2, \ldots, x_n\}$. This is a recursive definition: we already have a definition of list notation for sets when $n = 2$, and here we show how to define list notation when n has any value k greater than 2 on the assumption that we know how to define it when $n = k - 1$. Similarly, we define $\langle x_1, x_2 \ldots, x_n\rangle$ as $\langle x_1, \langle x_2, \ldots, x_n\rangle\rangle$.

Theorem: For any sets A and B, $A \cap B$ and $A - B$ exist. This is obvious from the forms of the definitions. If A is nonempty, $\bigcap A$ exists and the definition of the set does not depend on the choice of the element B.

Definition: $A \times B = \{x \in \mathcal{P}^{\mathbf{2}}(A \cup B) \mid (\exists a \in A.(\exists b \in B.x = \{\{a\}, \{a, b\}\}))\}$. We define A^2 as $A \times A$, and more generally define A^{n+1} as $A \times A^n$.

Theorem: $A \times B$ exists for all sets A and B. $A \times B = \{\langle a, b\rangle \mid a \in A \wedge b \in B\}$. The existence of $A \times B$ is obvious from the form of the definition. The trick is to notice that any pair $\langle a, b\rangle = \{\{a\}, \{a, b\}\}$ with $a \in A$ and $b \in B$ actually belongs to $\mathcal{P}^{\mathbf{2}}(A \cup B)$, because $\{a\}$ and $\{a, b\}$ both belong to $\mathcal{P}(A \cup B)$.

Definition: A *relation* is a set of ordered pairs. We define $x\,R\,y$ as $\langle x, y\rangle \in R$.

Observation: Just as in the type theory of chapter 2, not every logical relation is a set relation. For example, the logical relation of equality is not implemented as a set, because $Q = \{\langle x, y\rangle \mid x = y\}$ would have the unfortunate property $\bigcup^{\mathbf{2}} Q = V$, the universal set, which we know does not exist. For similar reasons, membership and inclusion are not set relations.

Definition: For any relation R, we define $\mathtt{fld}(R)$ as $\bigcup^{\mathbf{2}} R$, $\mathtt{dom}(R)$ as

$$\{x \in \mathtt{fld(R)} \mid (\exists y.\, \langle x, y\rangle \in R)\},$$

and $\mathtt{rng}(R)$ as

$$\{y \in \mathtt{fld(R)} \mid (\exists x.\, \langle x, y\rangle \in R)\}.$$

Theorem: The field, domain, and range of a relation R are sets. This is evident from the forms of the definitions. That they are the intended sets is evident from the fact that if $\langle x, y\rangle \in R$ then $x, y \in \bigcup^{\mathbf{2}} R$. Moreover, $\mathtt{fld}(R) = \mathtt{dom}(R) \bigcup \mathtt{rng}(R)$ and $R \subseteq \mathtt{dom}(R) \times \mathtt{rng}(R) \subseteq \mathtt{fld}(R)^2$.

† **Remark:** Once we have verified that we have an adequate foundation for the theory of relations, we can import definitions and concepts wholesale from type theory, always subject to the limitation that we cannot construct very large collections. For example we cannot define cardinals, ordinals, or general isomorphism types as equivalence classes under equinumerousness or isomorphism, because equinumerousness, isomorphism, and most of their equivalence classes are not sets. However, we can for example import every definition given in section 2.6, except that the collections $[=]$ and $[\subseteq]$, being too large, cannot be sets: however, for any set A, $[=] \cap A^2$ and $[\subseteq] \cap A^2$ are sets (one might not like definitions of these using the symbols $[=]$ or $[\subseteq]$, but the sets referred to can be defined as $\{\langle x, y\rangle \in A \times A \mid x = y\}$ and $\{\langle x, y\rangle \in A \times A \mid x \subseteq y\}$, respectively). This is a specific example of a general phenomenon: if R is a relation symbol, we cannot be sure that $[R] = \{\langle x, y\rangle \mid x\, R\, y\}$ exists, but for any sets A, B, we do know that $[R] \cap (A \times B) = \{\langle x, y\rangle \in A \times B \mid x\, R\, y\}$ exists.

3.2.1 Relations and Functions

We here reproduce the section on terminology for relations and functions with minor changes from the typed set theory chapter. Most but not all of

this is taken from section 2.6. The fact that few changes are needed makes an implicit point.

If A and B are sets, we define a *relation from A to B* as a subset of $A \times B$. A *relation* in general is simply a set of ordered pairs.

If R is a relation from A to B, we define $x\,R\,y$ as $\langle x, y\rangle \in R$. This notation should be viewed with care. In the superficially similar notations $x \in y$ or $x \subseteq y$, the symbols $\in, \subseteq$ do not denote sets at all: do not confuse logical relations with set relations.

If R is a relation, we define $\mathtt{dom}(R)$, the *domain of R*, as $\{x \in \mathtt{fld}(R) \mid (\exists y.x\,R\,y)\}$. We define R^{-1}, the *inverse of R*, as $\{\langle x, y\rangle \in \mathcal{P}^2(\bigcup^2 R) \mid y\,R\,x\}$. We define $\mathtt{rng}(R)$, the *range of R*, as $\mathtt{dom}(R^{-1})$. We note that $\mathtt{fld}(R)$, the *field of R*, already defined as $\bigcup^2 R$, is the union of $\mathtt{dom}(R)$ and $\mathtt{rng}(R)$. If R is a relation from A to B and S is a relation from B to C, we define $R|S$, the *relative product of R and S* as

$$\{\langle x, z\rangle \mid (\exists y.x\,R\,y \wedge y\,S\,z)\}.^5$$

The symbol $[=]\lceil A$ might be used to denote the equality relation restricted to A, the set $\{\langle x, x\rangle \in A \times A \mid x \in V\}$. Similarly $[\subseteq]\lceil\mathcal{P}(A)$ can be used as a name for a restriction of the subset relation, and so forth: the brackets convert a grammatical "transitive verb" to a noun.[6]

We define special characteristics of relations. Some of these terms are also used in connection with logical relations which do not determine sets: for example, the subset relation is reflexive, antisymmetric and transitive.

reflexive: R is *reflexive* iff $x\,R\,x$ for all $x \in \mathtt{fld}(R)$.

symmetric: R is *symmetric* iff for all x and y, $x\,R\,y \leftrightarrow y\,R\,x$.

antisymmetric: R is *antisymmetric* iff for all x, y if $x\,R\,y$ and $y\,R\,x$ then $x = y$.

asymmetric: R is *asymmetric* iff for all x, y if $x\,R\,y$ then $\neg y\,R\,x$. Note that this immediately implies $\neg x\,R\,x$.

transitive: R is *transitive* iff for all x, y, z if $x\,R\,y$ and $y\,R\,z$ then $x\,R\,z$.

[5] We leave it as an exercise for the reader to find a bound for the elements of this set, witnessing the fact that the set exists by Separation.

[6] The transformation of relation symbols into terms using brackets is an invention of ours and not likely to be found in other books.

equivalence relation: A relation is an *equivalence relation* iff it is reflexive, symmetric, and transitive.

partial order: A relation is a *partial order* iff it is reflexive, antisymmetric, and transitive.

strict partial order: A relation is a *strict partial order* iff it is asymmetric and transitive. Given a partial order R,

$$R - [=] = \{\langle x, y\rangle \mid x\, R\, y \wedge x \neq y\}$$

will be a strict partial order. From a strict partial order $R - [=]$, the partial order R can be recovered if it has no "isolated points" (elements of its field related only to themselves).

linear order: A partial order R is a *linear order* iff for any $x, y \in \mathtt{fld}(R)$, either $x\, R\, y$ or $y\, R\, x$. Note that a linear order is precisely determined by the corresponding strict partial order if its domain has two or more elements.

strict linear order: A strict partial order R is a *strict linear order* iff for any $x, y \in \mathtt{fld}(R)$, one has $x\, R\, y$, $y\, R\, x$ or $x = y$. If R is a linear order, $R - [=]$ is a strict linear order.

image: For any set $A \subseteq \mathtt{fld}(R)$, $R\text{“}A = \{b \mid (\exists a \in A.a\, R\, b)\}$.

extensional: A relation R is said to be *extensional* iff for any $x, y \in \mathtt{fld}(R)$, $R^{-1}\text{“}(\{x\}) = R^{-1}\text{“}(\{y\}) \rightarrow x = y$: elements of the field of R with the same preimage under R are equal. An extensional relation supports a representation of some of the subsets of its field by the elements of its field.

well-founded: A relation R is *well-founded* iff for each nonempty subset A of $\mathtt{fld}(R)$ there is $a \in A$ such that for no $b \in A$ do we have $b\, R\, a$ (we call this a minimal element of A with respect to R, though note that R is not necessarily an order relation).

well-ordering: A linear order R is a *well-ordering* iff the corresponding strict partial order $R - [=]$ is well-founded.

strict well-ordering: A strict linear order R is a *strict well-ordering* iff it is well-founded.

end extension: A relation S *end extends* a relation R iff $R \subseteq S$ and for any $x \in \mathtt{fld}(R)$, $R^{-1}\text{“}\{x\} = S^{-1}\text{“}\{x\}$. (This is a nonstandard adaptation of a piece of terminology from model theory).

function: f is a *function from A to B* (written $f : A \to B$) iff f is a relation from A to B and for all x, y, z, if $x\, f\, y$ and $x\, f\, z$ then $y = z$. For each $x \in \mathtt{dom}(f)$, we define $f(x)$ as the unique y such that $x\, f\, y$ (this exists because x is in the domain and is unique because f is a function). The notation $f[A]$ is common for the image $f\text{“}A$.

warning about function notation: Notations like $\mathcal{P}(x)$ for the power set of x should not be misconstrued as examples of the function value notation $f(x)$. There is no function $\mathcal{P}$ because the domain of such a function would be the collection of all sets, which cannot be a set in untyped set theory.

injection: A function f is an *injection* (or *one-to-one*) iff f^{-1} is a function.

surjection: A function f is a *surjection from A to B* or a *function from A onto B* iff it is a function from A to B and $f\text{“}A = B$.

bijection: A function f is a *bijection from A to B* iff it is an injection and also a surjection from A to B.

composition and restriction: If f is a function and A is a set (usually a subset of $\mathtt{dom}(f)$), define $f\restriction A$ as $f \cap (A \times V)$ (the *restriction of f to the set A*). If f and g are functions and $\mathtt{rng}(g) \subseteq \mathtt{dom}(f)$, define $f \circ g$ as $g|f$. This is called the *composition* of f and g. We may now and then write compositions as relative products, when the unnaturalness of the order of the composition operation is a problem.

identity function: Note that $[=]$ meets the specification of a function except that it fails to be a set. We call $[=]\restriction A$ the *identity function on A*, where A is any set: as the terminology suggests, each of these sets is a function.

abstraction: If $T[x]$ is a term (usually involving x) define $(x : A \mapsto T[x])$ or $(\lambda x : A.T[x])$ as $\{\langle x, T[x]\rangle \mid x \in A\}$. The explicit mention of the set A may be omitted when it is understood from context or from the form of the term $T[x]$.

isomorphism of relations: Relations R and S are said to be *isomorphic* iff there is a bijection f from $\mathtt{fld}(R)$ to $\mathtt{fld}(S)$ such that for every $x, y \in \mathtt{fld}(R)$ we have $x\,R\,y$ iff $f(x)\,S\,f(y)$. Such a bijection is called an *isomorphism from R to S*. Isomorphism between relations captures the idea that they have the same formal structure in certain sense. It is a valuable exercise to show that isomorphism is a (non-set) equivalence relation on set relations. In particular, if f is an isomorphism from R to S, then f^{-1} is an isomorphism from S to R.

terminology about partial orders: It is conventional when working with a particular partial order $\leq$ to use $<$ to denote $[\leq] - [=]$ (the corresponding strict partial order), $\geq$ to denote $[\leq]^{-1}$ (which is also a partial order) and $>$ to denote the strict partial order $[\geq] - [=]$.

A minimum of $\leq$ is an element m of $\mathtt{fld}(\leq)$ such that $m \leq x$ for all $x \in \mathtt{fld}(x)$. A maximum of $\leq$ is a minimum of $\geq$. A minimal element with respect to $\leq$ is an element m such that for no x is $x < m$. A maximal element with respect to $\leq$ is a minimal element with respect to $\geq$. Notice that a maximum or minimum is always unique if it exists. A minimum is always a minimal element. The converse is true for linear orders but not for partial orders in general.

For any partial order $\leq$ and $x \in \mathtt{fld}(\leq)$, we define $\mathtt{seg}_{\leq}(x)$ as $\{y \mid y < x\}$ (notice the use of the strict partial order) and $(\leq)_x$ as $[\leq] \cap (\mathtt{seg}_{\leq}(x))^2$. The first set is called the *segment* in $\leq$ determined by x and the second is called the *segment restriction* determined by x.

For any subset A of $\mathtt{fld}(\leq)$, we say that an element x of $\mathtt{fld}(\leq)$ is a lower bound for A in $\leq$ iff $x \leq a$ for all $a \in A$, and an upper bound for A in $\leq$ iff $a \leq x$ for all $a \in A$. If there is a lower bound x of A such that for every lower bound y of A, $y \leq x$, we call this the greatest lower bound of A, written $\inf_{\leq}(A)$, and if there is an upper bound x of A such that for all upper bounds y or A, we have $x \leq y$, we call this the least upper bound of A, written $\sup_{\leq}(A)$.

A special kind of partial order is a *tree*: a partial order $\leq_T$ with field T is a *tree* iff for each $x \in T$ the restriction of $\leq_T$ to $\mathtt{seg}_{\leq_T}(x)$ is a well-ordering. A subset of T which is maximal in the inclusion order among those well-ordered by $\leq_T$ is called a *branch*.

3.2.2 Exercises

1. Prove the lemma [used in class] that

$$(\forall xyzw.\{x,y\} = \{z,w\} \to ((x = z \wedge y = w) \vee (x = w \wedge y = z)).$$

2. Consider the original ordered pair definition of Wiener, $\langle x,y\rangle \equiv_{\texttt{def}} \{\{\{x\},\emptyset\},\{\{y\}\}\}$. Prove that this satisfies the basic properties needed for a notion of ordered pair to implement relations in set theory:

   (a)
   $$(\forall xyzw.\, \langle x,y\rangle = \langle z,w\rangle \to x = z \wedge y = w)$$

   (b) For any sets A, B, $A \times B = \{\langle x,y\rangle \mid x \in A \wedge y \in B\}$ exists.

   (c) For any set R of ordered pairs, $\texttt{dom}(R) = \{x \mid (\exists y.x\, R\, y)\}$ and $\texttt{rng}(R) = \{y \mid (\exists x.x\, R\, y)\}$ exist.

   Hint: think about how many members various sets involved in this definition have. By contrast, you cannot tell how many members $\{\{x\},\{x,y\}\}$ has, in general. Do you see why not?

3. Consider the ordered pair definition $\langle x,y\rangle \equiv_{\texttt{def}} \{\{x,0\},\{y,1\}\}$ [0 being defined as $\emptyset$ and 1 as $\{\emptyset\}$]. Prove that this satisfies the basic properties needed for a notion of ordered pair to implement relations in set theory:

   (a)
   $$(\forall xyzw.\, \langle x,y\rangle = \langle z,w\rangle \to x = z \wedge y = w)$$

   (b) For any sets A, B, $A \times B = \{\langle x,y\rangle \mid x \in A \wedge y \in B\}$ exists.

   (c) For any set R of ordered pairs, $\texttt{dom}(R) = \{x \mid (\exists y.x\, R\, y)\}$ and $\texttt{rng}(R) = \{y \mid (\exists x.x\, R\, y)\}$ exist.

4. Show that if we use our official definition of the ordered pair

$$\langle x,y\rangle \equiv_{\texttt{def}} \{\{x\},\{x,y\}\}$$

   that the theorem

$$(\forall xyzw.\, \langle x,y\rangle = \langle z,w\rangle \to x = z \wedge y = w)$$

   is true.

5. The natural number 1 was defined by Frege as the collection of all sets with exactly one element. Express "x has exactly one element" as a formula $\phi[x]$ using only propositional logic, quantifiers, equality and membership, and give a definition of the Frege natural number 1 in the form $\{x \mid \phi[x]\}$. Then prove that this set does not exist in Zermelo set theory. The first part of the question is readily answered by looking in chapter 2: it would be a better idea not to do this.

6. Show that if R and S are set relations (sets of ordered pairs), their relative product $R|S = \{\langle x, z\rangle \mid (\exists y.x\, R\, y \wedge y\, S\, z)\}$ exists, by finding a suitable set U such that $R|S = \{\langle x, z\rangle \in U \mid (\exists y, x\, R\, y \wedge y\, S\, z)\}$ (from which it follows that the set $R|S$ exists by Separation).

7. Prove directly that the set $\mathcal{Z}$ whose existence is asserted by the axiom of infinity is Zermelo-inductive. That is, prove that $\emptyset \in \mathcal{Z}$, then, assuming that $x \in \mathcal{Z}$, show that $\{x\} \in \mathcal{Z}$ must follow.

8. Is $\{\{x, 0\}, \{y, 1\}, \{z, 2\}\}$ (suppose that 2 is defined as $\{\{\emptyset\}\}$) an adequate definition of the ordered triple? Given an arbitrary set of this form, can we determine its first, second, and third component?

3.3 Case study: the implementation of the number systems in untyped set theory

In this section we will implement familiar systems of numbers in set theory. Part of the aim is to shed light on what it means to found mathematics on set theory. A general theme is that though we are identifying mathematical objects which we understand before studying set theory with certain sets, we are not really claiming to reveal anything about these objects, and moreover the identifications depend on decisions that could have been made differently: in some cases we will describe more than one alternative implementation of a concept, and we try to make it clear that choosing a different implementation would not change the underlying mathematics.

3.3.1 The natural numbers

We begin with the arithmetic of the natural numbers. Peano proposed a set of five axioms describing the arithmetic of the natural numbers in the nineteenth century[7]. It is worth noting that while these axioms do not impose an implementation of numbers themselves as sets, they do make essential use of sets. Later in the section we will give an alternative (and weaker) formulation not dependent on set theory at all.

Primitive notions: A constant 0, a unary operation σ (successor), and the set $\mathbb{N}$ of natural numbers.

Axiom 1: $0 \in \mathbb{N}$.

Axiom 2: For each $x \in \mathbb{N}$, $\sigma(x) \in \mathbb{N}$.

Axiom 3: For each $x \in \mathbb{N}$, $\sigma(x) \neq 0$.

Axiom 4: For each $x, y \in \mathbb{N}$, $\sigma(x) = \sigma(y) \rightarrow x = y$

Axiom 5: For each $S \subseteq \mathbb{N}$, if $0 \in S$ and $(\forall x \in S : \sigma(x) \in S)$, then $S = \mathbb{N}$.

We will give in this section and the following section not one but three implementations of Peano arithmetic. We will choose one of them as the

[7]We note that Peano's original axiom set used 1 as a primitive instead of 0.

official representation for our purposes, but we could equally well have chosen one of the others, and our mathematics would be essentially the same.

We review the definition of "inductive set" and the Axiom of Infinity.

Definition: A *Zermelo-inductive set* is defined as a set I such that $\emptyset \in I$ and $(\forall x \in I : \{x\} \in I)$.

Axiom of Infinity: We assert the existence of the set $\mathcal{Z}$ of all n such that for every Zermelo-inductive set I, n belongs to I.

First implementation of Peano arithmetic: We implement 0 as $\emptyset$, σ as $\{\langle x, \{x\}\rangle \in \mathcal{Z} \times \mathcal{P}(\mathcal{Z}) \mid x \in \mathcal{Z}\}$, and $\mathbb{N}$ as $\mathcal{Z}$. Notice the bounding of the definition of σ to verify that this set actually exists. To confirm that this is an implementation, we need to verify that the translations of each of the axioms hold:

Axiom 1: $\emptyset \in \mathcal{Z}$ holds because $\emptyset$ belongs to each Zermelo-inductive set, by the definition of "Zermelo-inductive", and to belong to each Zermelo-inductive set is to belong to $\mathcal{Z}$.

Axiom 2: We verify that for all $x \in \mathcal{Z}$, $\{x\} \in \mathcal{Z}$. Choose an $x \in \mathcal{Z}$ arbitrarily. Choose a Zermelo-inductive set I arbitrarily. Because $x \in \mathcal{Z}$, we have $x \in I$, by the definition of "Zermelo-inductive". Because I is Zermelo-inductive, we have $\{x\} \in I$. I was chosen arbitrarily, so we have that $\{x\}$ belongs to every Zermelo-inductive set, and so that $\{x\}$ belongs to $\mathcal{Z}$. The element $x \in \mathcal{Z}$ was chosen arbitrarily, so we have verified our claim.

Observation: The demonstrations of the interpreted Axioms 1 and 2 are together a direct proof that $\mathcal{Z}$ is itself Zermelo-inductive.

Axiom 3: For each $x \in \mathcal{Z}$, $\{x\} = \emptyset$ is obviously false, since the first set has an element and the second does not.

Axiom 4: We verify that for each $x, y \in \mathcal{Z}$, $\{x\} = \{y\} \rightarrow x = y$. Suppose $\{x\} = \{y\}$. Because $\{x\}$ is defined as $\{z \mid z = x\}$ (and exists by the Axiom of Elementary Sets) we have $x \in \{x\}$ (since $x = x$). Thus by substitution we have $x \in \{y\} = \{z \mid z = y\}$, so we have $x = y$. This is all quite obvious, but it is worth noting that such obvious things really can be proved.

Axiom 5: We need to verify that if $S \subseteq \mathcal{Z}$ and $\emptyset \in S$ and $(\forall x \in S : \{x\} \in S)$, it follows that $S = \mathcal{Z}$. This is very direct: the conditions imply immediately that S is Zermelo-inductive, whence it follows that $\mathcal{Z} \subseteq \mathcal{S}$ ($\mathcal{Z}$ is a subset of any Zermelo-inductive set, since an element of $\mathcal{Z}$ belongs to all Zermelo-inductive sets and so to that specific one), whence it follows that $S = \mathcal{Z}$ by Extensionality (if $A \subseteq B$ and $B \subseteq A$, then A and B are sets with the same elements and so are equal).

Because of the possibility of this interpretation, we may refer to elements of $\mathcal{Z}$ as "Zermelo natural numbers". This is not our official interpretation, but it is not a bad one, and we will indicate in this section and the next one how we would proceed if we chose to use it as our official implementation.

Our official interpretation relies on a different choice of implementation of the successor operation, and actually on a different form of the Axiom of Infinity, which turns out not to be provable in Zermelo's original theory.

We reformulate the definition of "inductive set" and the Axiom of Infinity.

Definition: For any set x, we define x^+ as $x \cup \{x\}$.

Definition: A *von Neumann-inductive set* is defined as a set I such that $\emptyset \in I$ and $(\forall x \in I : x^+ \in I)$.

Axiom of Infinity*: We assert the existence of the set N of all n such that for every von Neumann-inductive set I, n belongs to I.

Second (and official) implementation of Peano arithmetic: We implement 0 as $\emptyset$, σ as $\{\langle x, x \cup \{x\}\rangle \in N \times \mathcal{P}(N)) \mid x \in N\}$, and $\mathbb{N}$ as N. Notice the bounding of the definition of σ to verify that this set actually exists. To confirm that this is an implementation, we need to verify that the translations of each of the axioms hold:

Axiom 1: $\emptyset \in N$ holds because $\emptyset$ belongs to each von Neumann-inductive set, by the definition of "von Neumann-inductive", and to belong to each von Neumann-inductive set is to belong to N.

Axiom 2: We verify that for all $x \in N$, $x^+ \in N$. Choose an $x \in N$ arbitrarily. Choose a von Neumann-inductive set I arbitrarily. Because $x \in N$, we have $x \in I$, by the definition of "von Neumann-inductive". Because I is von Neumann-inductive, we have $x^+ \in I$.

I was chosen arbitrarily, so we have that x^+ belongs to every von Neumann-inductive set, and so that x^+ belongs to N. The element $x \in N$ was chosen arbitrarily, so we have verified our claim.

Observation: The demonstrations of the interpreted Axioms 1 and 2 are together a direct proof that N is itself von Neumann-inductive.

Axiom 3: For each $x \in N$, $x^+ = \emptyset$ is obviously false, since the first set has an element and the second does not.

Axiom 4: We want to verify that for each $x, y \in \mathcal{Z}$, $x^+ = y^+ \rightarrow x = y$. We will ask the reader to prove this, with some guidance, in an exercise.

Axiom 5: We need to verify that if $S \subseteq N$ and $\emptyset \in S$ and $(\forall x \in S : x^+ \in S)$, it follows that $S = N$. This is very direct: the conditions imply immediately that S is von Neumann-inductive, whence it follows that $N \subseteq S$ (N is a subset of any von Neumann-inductive set, since an element of N belongs to all von Neumann-inductive sets and so to that specific one), whence it follows that $S = N$ by Extensionality (if $A \subseteq B$ and $B \subseteq A$, then A and B are sets with the same elements and so are equal).

We make an important observation at this point. As long as our implementation has the characteristic that $\mathbb{N}$ is defined as the intersection of all sets I which contain 0 and satisfy $(\forall x \in I : \sigma(x) \in I)$, and we can verify that this set exists, the verification of Axioms 1, 2, and 5 goes exactly as above. Only the verifications of Axioms 3 and 4 will depend on the details of what object is chosen to implement 0 and what operation is chosen to implement σ. The reader can confirm this by reading the parallel demonstrations of Axioms 1, 2, and 5 given in the two implementations given so far, which do not depend in any way on any specific information about 0 or σ.

A major practical advantage of the von Neumann representation is that the von Neumann implementation of each natural number n is $\{0, \ldots, n-1\}$, a set with n elements, which facilitates reasoning about counting (discussed in the next subsection). A further and more profound advantage is that this representation generalizes naturally to a representation of transfinite ordinals, which is not the case for the Zermelo representation.

We claim to have a representation of the natural numbers at this point, but the reader may notice that we have not defined even such basic concepts as addition and multiplication. We proceed to repair this lack.

Iteration Theorem: For each set A and function $f : A \to A$, and element $a \in A$, there is a unique function $g : \mathbb{N} \to A$ such that $g(0) = a$ and for each $n \in \mathbb{N}$, $g(\sigma(n)) = f(g(n))$. Once the theorem is proved, we introduce the notation $\texttt{iter}_{f,a}$ for the unique function g and the notation $f^n(a)$ for $g(n)$ [this last may also serve to make our motivation clear].

Proof of the Iteration Theorem: Fix a set A, a function $f : A \to A$, and an element $a \in A$.

Definition: We define an (f, a)-inductive set as a set $I \subseteq \mathbb{N} \times A$ such that $\langle 0, a)\rangle \in I$ and for all $\langle n, x\rangle \in I$, we also have $\langle \sigma(n), f(a)\rangle \in I$. Further define g as the set of all elements of $\mathbb{N} \times A$ which belong to all (f, a)-inductive sets.

We first prove that g is a function, that is, for each $n \in \mathbb{N}$, there is exactly one $x \in A$ such that $\langle n, x\rangle \in g$. We prove this by induction on n.

basis: We claim that there is exactly one x such that $\langle 0, x\rangle \in g$. We claim in fact that $x = 0$. We know that $\langle 0, a\rangle \in g$, because $\langle 0, a\rangle$ belongs to every (f, a)-inductive set, and that is the criterion for belonging to g. Now suppose that $y \neq 0$; our aim is to show that $\langle 0, y\rangle \neq g$. Let I be an (f, a)-inductive set: we claim that $J = I - \{\langle 0, y\rangle\}$ is also (f, a)-inductive. Certainly $\langle 0, a\rangle \in J$, since $\langle 0, a\rangle \in I$ and $\langle 0, a\rangle \neq \langle 0, y\rangle$. Suppose $\langle n, z\rangle \in J$. It follows that $\langle \sigma(n), f(z)\rangle \in I$, because I is (f, a)-inductive and $J \subseteq I$; but also $\langle \sigma(n), f(z)\rangle \neq \langle 0, y\rangle$ by Axiom 3, so $\langle \sigma(n), f(z)\rangle \in J$, so J is (f, a)-inductive, so $\langle 0, y\rangle \notin g$, since $\langle 0, y\rangle \notin J$, an (f, a)-inductive set. This completes the proof of the basis.

induction step: We assume for a fixed $k \in \mathbb{N}$ that there is exactly one x such that $\langle k, x\rangle \in g$, and show that there is exactly one y such that $\langle \sigma(k), y\rangle \in g$. There is at least one such y, namely $f(x)$, because $\langle \sigma(k), f(x)\rangle \in g$, since each (f, a)-inductive set contains $\langle k, x\rangle \in g$, and so contains $\langle \sigma(k), f(x)\rangle$. Now consider any $z \neq f(x)$: our aim is to show $\langle \sigma(k), z\rangle \notin g$. Let I be any (f, a)-inductive set: we show that $J = I - \{\langle \sigma(k), z\rangle\}$ is also (f, a)-inductive. $\langle 0, a\rangle \in I$ of course, and $\langle 0, a\rangle \neq \langle \sigma(k), z\rangle$ by Axiom 3, so $\langle 0, a\rangle \in J$. Now suppose that $\langle n, w\rangle \in J$: certainly

$\langle \sigma(n), f(w) \rangle \in I$, but further $\langle \sigma(n), f(w) \rangle \neq \langle \sigma(k), z \rangle$, because if this equation held we would have $n = k$ by Axiom 4, and we know that if $n = k$ we have $w = x$, so $z = f(x)$, which contradicts our choice of z. And thus $\langle \sigma(n), f(w) \rangle \in J$, so J is (f, a)-inductive, whence $\langle \sigma(k), z \rangle \notin J$ cannot belong to g, which completes the proof of the induction step.

Since g is a function, we can now see that $g(0) = a$ and $g(\sigma(n)) = f(g(n))$. We further claim that for any function g' such that $g'(0) = a$ and $g'(\sigma(n)) = f(g'(n))$, we have $g'(n) = g(n)$ for each $n \in \mathbb{N}$, whence $g = g'$, establishing uniqueness. This is an easy induction. $g(0) = a = g'(0)$ is obvious. If $g(k) = g'(k)$, then $g(\sigma(k)) = f(g(k)) = f(g'(k)) = g'(\sigma(k))$.

This completes the proof of the Iteration Theorem.

We state some identities for the iteration notation $f^n(a)$. Notice that, where g is the unique function provided by the Iteration Theorem, $f^0(a) = g(0) = a$, so we obtain the identity $f^0(a) = a$. We further obtain $f^{\sigma(n)}(a) = g(\sigma(n)) = f(g(n)) = f(f^n(a))$, so we have the identity $f^{\sigma(n)}(a) = f(f^n(a))$. It is also worth noting that we can define the function f^n as

$$\{\langle x, y \rangle \in A \times A \mid y = f^n(x)\},$$

and this will define a function, even though the notation $f^n(a)$ was not originally defined as a function application notation; it is safe to read it that way, anyway. In English, $f^n(x)$ is defined as $\texttt{iter}_{f,x}(n)$, recalling that $\texttt{iter}_{f,x}$ is the unique function g from $\mathbb{N}$ to A provided by the Iteration Theorem for which $g_x(0) = x; g_x(\sigma(n)) = f(g_x(n))$, for all $n \in \mathbb{N}$. Something rather subtle is going on here: f^n is a function whose value at each $x \in A$ is determined by applying a function depending on x to n.

We can now define some familiar operations.

definition of addition of natural numbers: For natural numbers m, n, we define $m + n$ as $\sigma^n(m)$. Note that if we define 1 as $\sigma(0)$, we can represent $\sigma(n)$ in the more familiar form $n + 1$.

definition of multiplication of natural numbers: For each natural number n, define $\texttt{add}_n$ as $\{\langle m, m + n \rangle \in \mathbb{N}^2 \mid m \in \mathbb{N}\}$, the function which adds n. Define $m \cdot n$ as $(\texttt{add}_m)^n(0)$.

arithmetic rules from the iteration theorem: We derive rules for addition and multiplication which are given as additional axioms for Peano arithmetic when it is formulated independently of set theory.

1. $n + 0 = \sigma^0(n) = n$
2. $m + \sigma(n) = \sigma^{\sigma(n)}(m) = \sigma(\sigma^n(m)) = \sigma(m + n)$
3. $n \cdot 0 = (\mathtt{add}_n)^0(0) = 0$
4. $$m{\cdot}\sigma(n) = (\mathtt{add}_m)^{\sigma(n)}(0) = \mathtt{add}_m((\mathtt{add}_m)^n(0)) = \mathtt{add}_m(m{\cdot}n) = m{\cdot}n{+}m$$

The formulation of Peano arithmetic independently of set theory adds addition and multiplication as new primitive notions (with closure properties added as part of axiom 2), takes the equations proved just above as Axioms 6-9, and modifies Axiom 5 to assert for any formula $\phi(x)$ for which $\phi(0)$ is true and $(\forall k \in \mathbb{N} : \phi(k) \to \phi(\sigma(k)))$ is true, we obtain $(\forall n \in \mathbb{N} : \phi(n))$. We will not make use of this more restricted formulation, since we have no reason to refrain from using set concepts. For us, "Axioms" 6-9 are consequences of Axioms 1-5 with Axiom 5 in its original form involving sets rather than open sentences.

More general forms of recursion can be implemented, and indeed our Iteration Theorem is a somewhat unusual formulation.

Recursion Theorem: Let A be any set, let $a \in A$, and let $f : \mathbb{N} \times A \to A$. Then there is a unique function g such that $g(0) = a$ and for all $n \in \mathbb{N}$, we have $g(n + 1) = f(n, g(n))$.

Proof: Define $F(\langle n, x\rangle) = \langle n + 1, f(n, g(n))\rangle$. Then the function g is definable using the Iteration Theorem as $g(n) = \pi_2(F^n(\langle 0, a\rangle))$.

We give an example to illustrate yet more complex forms of recursion.

Recursion example (Fibonacci numbers): Define $f(\mathbb{N}) \times \mathbb{N} \to \mathbb{N}$ by $f(\langle m, n\rangle) = \langle n, m + n\rangle$. Then the nth Fibonacci number $F(n)$ can be defined as $\pi_1(F^n(\langle 1, 1\rangle))$.

We introduce an even stronger form of recursion.

Theorem (course of values recursion): Let $\mathcal{F}$ be the set of all functions whose domain is a set $\{m \in \mathbb{N} \mid m < n\}$ for $n \in \mathbb{N}$ (of course, if we use the von Neumann definition of the natural numbers, this domain is n itself) and whose range is A. Let $F : \mathcal{F} \to A$. Then there is a uniquely determined function $G : \mathbb{N} \to A$ such that

$$G(n) = F(G\lceil\{m \in \mathbb{N} \mid m < n\}).$$

for each natural number n.

Proof of course-of-values recursion theorem: Define $H : \mathcal{F} \to \mathcal{F}$ as follows: if f has domain $\{m \in \mathbb{N} \mid m < n\}$, define $H(f)$ as $f \cup \{\langle n, F(f)\rangle\}$. Now define G as $\bigcup\{H^n(\emptyset) \in \mathcal{P}(\mathcal{F}^2) \mid n \in \mathbb{N}\}$.

The reader still may be suspicious of our claim that this is an adequate axiomatization of arithmetic. We refer him or her to section 2.8 in the previous chapter, in which various basic results of arithmetic are proved from the Peano axioms (including 6-9), the formal differences being the convention that all quantifiers are taken to be restricted to the natural numbers, since natural numbers are the only objects we are talking about (so there is no reason to write $(\forall k \in \mathbb{N} : \phi)$: this is abbreviated to $(\forall k : \phi)$).

We have another, quite abstract point, to make about implementations. We are planning to use the von Neumann implementation, under which $0 = \emptyset, \sigma(x) = x^+$, and $\mathbb{N}$ is the set N of von Neumann natural numbers. We want to argue that if we are careful, we can translate any theorem proved in set theory (not just arithmetic) which mentions natural numbers (using the von Neumann interpretation) into a theorem proved in set theory which mentions natural numbers using any other implementation we might choose. The idea is that we can eliminate all consideration of which sets the natural numbers are, and rely on no fact about the natural numbers other than which one is zero and which natural numbers are successors of which other natural numbers. The key point to observe is that if n is a natural number, $x \in n$ is equivalent to $x < n$, which we define (as we did in section 2.8 above) as $(\exists k \in \mathbb{N}.x + k = n) \wedge x \neq n$. Set theory allows us to deduce that $x \in 0$ is false and that $x \in \sigma(n) = n \cup \{n\}$ iff $x \in n \vee x = n$. We proved in section 2.8 that $x < 0$ is false and that $x \leq n \leftrightarrow x < n + 1$, which is equivalent to $x < \sigma(n) \leftrightarrow x < n \vee x = n$. Set theory allows us to deduce that $(\forall k \in \mathbb{N} : k \in m \leftrightarrow k \in n) \to m = n$, by applying Extensionality. The assertion $(\forall k \in \mathbb{N} : k < m \leftrightarrow k < n) \to m = n$ follows from the fact

that $\leq$ is a linear order, proved in section 2.8: suppose that $m \neq n$; then either $m \leq n$ or $n \leq m$; suppose $m \leq n$ without loss of generality; then $m < n$, but of course $\neg m < m$, so $(\forall k \in \mathbb{N} : k < m \leftrightarrow k < n)$ does not hold. To make the statement (and proof) of a theorem involving natural numbers implementation-independent, make sure that all references to what objects are elements of a natural number n are replaced with references to which objects are natural numbers less than n (a concept which can be defined independently of the choice of implementation).

Exercises

1. Prove Axiom 4 for the von Neumann implementation: that is, prove that for any von Neumann natural numbers x, y, $x^+ = y^+ \rightarrow x = y$.

   First prove this using the special assumption that for all sets x, y, it is not the case both that $x \in y$ and $y \in x$. You should find that this makes it possible to prove that $x^+ = y^+ \rightarrow x = y$ for all sets x, y, and so of course for all natural numbers x, y. [This special assumption is a consequence of the Axiom of Foundation, which we will introduce and add to our official set theory later, but which we do not have yet.]

   This should give a hint about how to prove the result for all natural numbers without the special assumption: demonstrate by induction that for all natural numbers x, y, it is not the case that both $x \in y$ and $y \in x$ hold. Prove this theorem, then show how to use it to prove the von Neumann version of Axiom 4 without making any special assumptions.

   A further hint: a set x is said to be *transitive* iff for all $y \in x$, for all $z \in y$, we also have $z \in x$ (a transitive set contains all elements of its own elements). It is useful to prove (by induction of course) that all von Neumann natural numbers are transitive.

   The moral here is that if we have Foundation (which we will have), Axiom 4 for our implementation is almost as easy as it is for the Zermelo implementation. But it is useful to see that the validity of the von Neumann interpretation does not depend on assuming Foundation as an axiom.

2. Prove by mathematical induction that for all natural numbers n, $\sigma^n(0) = n$. (the notation $\sigma^n(0)$ is a special case of the notation $f^n(a)$ defined in

the Iteration Theorem). You might want to look at the identities for the notation $f^n(a)$ which I supply in a new paragraph after the proof of the Iteration Theorem. This isn't precisely hard, but you have to pay attention to what you write.

3. Prove the associative law of addition as a theorem of our set theory. Prove it quite straightforwardly using more or less the same strategy I used to prove commutativity of addition in section 2.8 (or in class): be aware that you have to prove it by induction, and set up the appropriate basis step and induction step and prove them using axioms 6 and 7, proving further statements by induction if necessary.

3.3.2 The natural numbers and counting elements of sets

In this section, we will discuss the original use of the natural numbers for counting finite sets.

Definition: Fix a set X. We define a sequence of subsets of $\mathcal{P}(X)$. Define $[X]^0$ as $\{\emptyset\}$. Define $[X]^{n+1}$ as $\{u \cup \{x\} \in \mathcal{P}(X) \mid u \in [X]^n \wedge x \in X \setminus u\}$.

There is a function K such that $K(n) = [X]^n$ by the Iteration Theorem, with $A = \mathcal{P}^2(X), a = \{\emptyset\}$, and f defined by

$$f(U) = \{u \cup \{x\} \mid u \in U \wedge x \in X - u\}$$

for $U \in \mathcal{P}^2(X)$.

The natural reading of $[X]^n$ is "the collection of subsets of X with n elements".

Definition: The collection $[X]^{<\omega}$ of finite subsets of X is defined as the intersection of all sets $I \subseteq \mathcal{P}(X)$ with the property that $\emptyset \in I$ and for every $u \in I, x \in X$, $u \cup \{x\} \in I$. A set I with this property is said to be "X-finite-inductive". The natural reading of $[X]^{<\omega}$ is "the collection of finite subsets of X", and we say that a set is a finite subset of X iff it belongs to $[X]^{<\omega}$. We say that a set X is finite iff $X \in [X]^{<\omega}$, and infinite iff it is not finite.

Observation: It is provable that for any $A \subseteq X$, $A \in [X]^{<\omega}$ iff

$$(\exists n \in \mathbb{N} \mid A \in [X]^n\}.$$

Prove that $\bigcup\{[X]^n \mid n \in \mathbb{N}\}$ satisfies the right closure property to show that $[X]^{<\omega} \subseteq \bigcup\{[X]^n \mid n \in \mathbb{N}\}$, then prove by mathematical induction on n that each $[X]^n \subseteq [X]^{<\omega}$.

We introduce an important concept of set theory whose possibilities we will not be exhausting in this section.

Definition: Sets A and B are said to be *equinumerous* or less formally, to be of the same cardinality, or, even less formally, to be of the same size, iff there is a bijection $f : A \to B$ from A onto B.

Theorem: The relation $\sim$ is reflexive, symmetric, and transitive: for any set A, the restriction of $\sim$ to $\mathcal{P}(A)^2$ is an equivalence relation. (The qualification of the claim that $\sim$ is an equivalence relation is that it is not a set).

Proof of Theorem: The identity relation is a bijection from A to A, so $A \sim A$. If $A \sim B$, then there is $f : A \sim B$, a bijection from A to B, and $f^{-1} : B \to A$ is a bijection from B to A, so $B \sim A$ as well. If $A \sim B$ and $B \sim C$, then there is $f : A \to B$, a bijection from A onto B, and $g : B \to C$, a bijection from B onto C, and $g \circ f$ is a bijection from A onto C, so $A \sim C$ as well.

Finite Counting Theorem: If $A, B \in [X]^{<\omega}$, then

$$A \sim B \leftrightarrow (\exists n \in \mathbb{N} : A \in [X]^n \wedge B \in [X]^n),$$

and further $[X]^m \cap [X]^n = \emptyset$ when $m \neq n$.

Proof of Theorem: We have already shown that for any $A \in [X]^{<\omega}$ there is n such that $A \in [X]^n$.

We prove by induction on n that if $A \in [X]^n$ and $B \in \mathcal{P}(X)$, then $A \sim B \leftrightarrow B \in [X]^n$.

For all $A \in [X]^0$ and for all subsets B of X, we immediately have $A \sim B$ iff B is empty, that is, B is also in $[X]^0$.

Now suppose that for any $A \in [X]^k$, we have, for all B subsets of X, that $A \sim B$ iff $B \in [X]^k$. Suppose that $C \in [X]^{k+1}$ and $D \in \mathcal{P}(X)$. Our aim is to show that $C \sim D \leftrightarrow D \in [X]^{k+1}$. $C = E \cup \{x\}$ for some $E \in [X]^k$ and $x \notin E$. If $C \sim D$ is witnessed by a bijection f, then $D = f\text{“}C = f\text{“}E \cup \{f(x)\}$. By ind hyp $f\text{“}E$ belongs to $[X]^k$ and obviously $f(x) \notin f\text{“}E$, so $D \in [X]^{k+1}$. Now suppose that $D \in [X]^{k+1}$, so $D = F \cup \{y\}$ where $F \in [X]^k$ and $y \notin F$. By ind hyp, $E \sim F$, and if $g : E \to F$ is a bijection onto F witnessing this, then $g \cup \{\langle x, y\rangle\}$ is a bijection witnessing $C \sim D$.

This completes the proof by induction of the first claim.

Now we show by induction on n that for all $A \in [X]^n$ and $m \neq n$ a natural number, we have $A \notin [X]^m$.

This is evident for $n = 0$, as if $A \in [X]^n$ we have A empty, and for any $m \neq 0$, all elements of $[X]^m$ are nonempty.

Fix $k \in \mathbb{N}$, and assume that for all $A \in [X]^k$ and $m \neq k$, $A \notin [X]^m$. Suppose that $A \in [X]^{k+1}$ and also $A \in [X]^p$ with $p \neq k+1$ (our aim is a contradiction). We know that $p \neq 0$, so $p = m+1$ for some m. We have $A = B \cup \{x\}$ for some $B \in [X]^k$ and $x \notin B$, and $A = C \cup \{y\}$ for some $C \in [X]^m$ and $y \notin C$. Now if $x = y$ we have $B = C$ which contradicts the induction hypothesis. If $x \neq y$, we observe that $\mathtt{id}_A \setminus \{\langle y, y\rangle\} \cup \{\langle y, x\rangle\}$ is a bijection from B to C, so $B \sim C$, which contradicts the inductive hypothesis. In both cases, as soon as we know $B \sim C$, we know that $C \in [X]^k$ by the first claim, and then by inductive hypothesis we know that $[X]^k$ is the only set of this form to which C belongs, so $k = m$, which is a contradiction.

This completes the proof of the second claim and the entire theorem.

We present a further formulation of the Axiom of Infinity and implementation of the natural numbers in Zermelo set theory, which we will not use, but which we offer for comparison with the treatment of the natural numbers in chapter 2.

Axiom of Infinity**: There is a set $\mathcal{I}$ such that $\mathcal{I} \notin [\mathcal{I}]^{<\omega}$: i.e., $\mathcal{I}$ is not finite.

The status of this version of the Axiom of Infinity is that it is weaker than (that is, a consequence of) either of the other forms we have given (neither of which is deducible from the other). Neither of the other axioms can be proved from this one.

An alternative implementation of the natural numbers: We give an alternative implementation of the natural numbers, based directly on counting sets, like the implementation in chapter 2.

zero: Define 0 as $\{\emptyset\}$.

successor: Define $\sigma(n)$ (for $n \in \mathcal{P}^2(\mathcal{I})$) as

$$\{u \cup \{x\} \in \mathcal{P}(\mathcal{I}) \mid u \in n \wedge x \notin u\}.$$

$\mathcal{I}$-inductive: We say that a set I is $\mathcal{I}$-inductive iff $0 \in I$ and $(\forall n : n \in I \to \sigma(n) \in I)$.

definition of $\mathcal{N}$: We define $\mathcal{N}$ as the collection of all elements of $\mathcal{P}^2(\mathcal{I})$ which belong to all $\mathcal{I}$-inductive sets.

axioms 1,2, and 5: Verification of these axioms goes exactly as in the other two implementations we have discussed.

axiom 3: We need to show that for any $n \in \mathcal{N}$, we have $\sigma(n) \neq 0$. This is straightforward: each element of $\sigma(n)$ is of the form $u \cup \{x\}$ and so must have an element, and 0 has $\emptyset$ as an element, which does not have any elements.

axiom 4: We need to show that for any $m, n \in \mathcal{N}$, if $\sigma(m) = \sigma(n)$ then $m = n$. This requires an extended argument.

We notice first that for each $n \in \mathbb{N}$, for each $u \in n$, we have $u \in [\mathcal{I}]^{<\omega}$. The reason for this is that if we take any $\mathcal{I}$-finite-inductive set (a set $I \subseteq \mathcal{P}(\mathcal{I})$ such that $\emptyset \in I$ and

$$(\forall ux : u \in I \wedge x \in \mathcal{I} \rightarrow u \cup \{x\} \in I))$$

we can prove by an obvious direct induction that each natural number n is a subset of I.

It then follows that no natural number is empty. 0 is nonempty. Suppose that n is nonempty and $u \in n$. We have shown that $u \in [\mathcal{I}]^{<\omega}$, and we know that $\mathcal{I} \notin [\mathcal{I}]^{<\omega}$ and that u is a subset of $\mathcal{I}$, so u is a proper subset of $\mathcal{I}$, there is $x \in \mathcal{I} - u$, and $u \cup \{x\} \in \sigma(n)$, so $\sigma(n)$ is nonempty, and all elements of $\mathbb{N}$ are nonempty sets by induction.

We prove a lemma. For each n, if $u \in \sigma(n)$ and $x \in u$, then $u - \{x\} \in n$ (taking an element away from an $n+1$-element set gives an n-element set). If $u \in [\mathcal{I}]^1$ and $x \in u$, then u is $\emptyset \cup \{y\}$ for some y, so $u = \{y\}$, whence $y = x$, so $u \setminus \{x\} = \emptyset \in [\mathcal{I}]^0$. Now fix $k \in \mathbb{N}$ and suppose that for every $u \in [\mathcal{I}]^{k+1}$ and $x \in u$ we have $u \setminus \{x\} \in [\mathcal{I}]^k$. Suppose that $u \in [\mathcal{I}]^{k+2}$ and $x \notin u$. Our aim is to show that $u - \{x\} \in [\mathcal{I}]^{k+1}$. Because $u \in [\mathcal{I}]^{k+2}$, u is of the form $v \cup \{y, z\}$ where $v \in [\mathcal{I}]^k$, $y \notin v$, $v \cup \{y\} \in [\mathcal{I}]^{k+1}$, and $z \notin v \cup \{x\}$. If $x = z$, we have $u \setminus \{z\} = v \cup \{y\} \in [\mathcal{I}]^{k+1}$ as desired. Otherwise we have $x \in v \cup \{y\}$, and by inductive hypothesis we have $(v \cup \{y\}) \setminus \{x\} \in [\mathcal{I}]^k$, and since $z \notin v \cup \{y\} \setminus \{x\}$, we have $((v \cup \{y\}) \setminus \{x\}) \cup \{z\} = u \setminus \{x\} \in [\mathcal{I}]^{k+1}$ as required. This completes the proof by induction of the lemma.

Now suppose that $\sigma(m) = \sigma(n)$. Both m and n are nonempty. An element of $\sigma(m)$ must be of the form $u \cup \{x\}$ where $u \in [\mathcal{I}]^m$

and $x \notin u$, and also of the form $v \cup \{y\}$ where $v \in [\mathcal{I}]^n$ and $y \notin v$. Now if $x = y$ we have $u = v$ and $m = n$ immediately. If $x \neq y$, observe that since $u \cup \{x\}$ is in $[\mathcal{I}]^{m+1}$ and $y \in u$, we must have $v = (u \cup \{x\}) \setminus \{y\} \in [\mathcal{I}]^m$, whence v is in both $[\mathcal{I}]^m$ and $[\mathcal{I}]^m$, whence $m = n$. This completes the verification of Axiom 4.

If one assumes either the Zermelo or von Neumann form of the Axiom of Infinity, it is easier to prove that the representation just given works. One first proves that $\mathcal{Z}$ or N is an infinite set (in the case of N, this is easy, as $n \in [N]^n$ is straightforward to prove). One then observes that the new representation of n is $[\mathcal{I}]^n$ (where the superscript represents the previous implementation of n). Axiom 3 remains easy. Axiom 4 holds because if $[\mathcal{I}]^{m+1} = [\mathcal{I}]^{n+1}$, then $m + 1 = n + 1$ by the Finite Counting Theorem, so $m = n$ by Axiom 4 for the original implementation of the natural numbers, whence of course $[\mathcal{I}]^m = [\mathcal{I}]^n$.

We are interested in this representation because of its evident relationship to the implementation given in chapter 2. It is also interesting to present a representation of the natural numbers in Zermelo set theory which does not actually rely on explicitly postulating a zero set and a successor operation, but just assumes in the abstract that there is an infinite set. It is worth noticing that the existence of *any* implementation of the natural numbers implies that the representation of n as $[\mathcal{I}]^n$ works; to show this we need to demonstrate that $\mathbb{N}$ is an infinite set no matter how the natural numbers are implemented, a proof of which is suggested as an exercise. So this is demonstrably the weakest form of the Axiom of Infinity for Zermelo set theory.[8]

Now addition and multiplication can be defined in natural ways that we learned in elementary school (and which generalize to addition and multiplication operations on infinite cardinals).

Definition: The cardinality $|A|$ of a finite set A is defined as the natural number n such that $A \in [A]^n$. That this definition makes sense for each finite set is established by results proved above.

[8] A lacuna in this last representation is that we are told nothing about the identity of the postulated infinite set $\mathcal{I}$. We put an evil suggestion on the table (which contradicts the axiom of foundation): what if $\mathcal{I} = \mathbb{N}$ (this last being understood in terms of the third implementation)? This can be made to work but is admittedly quite weird.

Since we are using the von Neumann representation, we can further observe that $|A|$ is the unique natural number such that $A \sim n$, though this requires some additional care. A straightforward induction shows that for each $n \in \mathbb{N}$, we have $n \in [n]^n$: each von Neumann natural number n is an n-element subset of itself.

We need the further lemma that $X \subseteq Y \rightarrow [X]^n \subseteq [Y]^n$. A set A belongs to $[X]^0$ iff it belongs to $[Y^0]$. Suppose for a fixed k that for any set B, B belongs to $[Y]^k$ if it belongs to $[X]^k$. Suppose that $A \in [X]^{k+1}$. There must be $x \in X$ such that $A = B \cup \{x\}$, $x \notin B$, and $B \in [X]^k$ and so $B \in [Y]^k$ by ind hyp. But then $x \in Y$ and $x \notin B$ so $A = B \cup \{x\} \in [Y]^{k+1}$.

Then we can argue that if $A \in [A]^n$, we can further argue using the lemma that both A and n belong to $[A \cup n]^n$, whence we have $A \sim n$. Conversely, if $A \sim n$ we want to argue that $A \in [A]^n$: this seems best proved by induction; if $A \sim 0$, then A is the empty set and $A \in [A]^0$; if for a fixed k, for any set B, $B \sim k$ implies $B \in [B]^k$, consider an arbitrary A such that $A \sim k+1$. If f is a bijection from A onto $k+1$, $f^{-1}\text{“}k$ is equinumerous to k and so $f^{-1}\text{“}k \in [f^{-1}\text{“}k]^k$ by ind hyp, whence $f^{-1}\text{“}k \in [A]^k$ by an earlier lemma, whence $A = f^{-1}\text{“}k \cup \{f^{-1}(k)\} \in [A]^{k+1}$ as desired.

Definition: Suppose $|A| = m$ and $|B| = n$. Define $m \oplus n$ as

$$|(A \times \{0\}) \cup (B \times \{1\})|,$$

and define $m \otimes n$ as $|A \times B|$. We further state our eventual intention to define $|A| + |B|$ as

$$|(A \times \{0\}) \cup (B \times \{1\})|$$

and $|A| \cdot |B|$ as $|A \times B|$ for all sets A, B, when we have a more general notion of cardinality (under which $|A| = |B| \leftrightarrow A \sim B$ will hold for all sets A, B).

These meanings are assigned to $\oplus$ and $\otimes$ only in this section; the intention is to show that they are the same as $+$ and $\cdot$.

Some theorems need to be proved to verify that the last definition makes sense, and then of course we want to prove that $m \oplus n = m + n$ and $m \otimes n = m \cdot n$.

The definition given for addition can be given a more abstract form:

Abstract definition of addition: Suppose that $|A| = m$ and $|B| = n$, and that $A \cap B = \emptyset$. Define $m \oplus n$ as $|A \cup B|$. (This could also be stated for general sets when we have a general definition of cardinality).

To verify that this makes sense, we need to show that if $|A| = |A'| = m$ and $|B| = |B'| = n$, and $A \cap B = A' \cap B' = \emptyset$, it follows that $|A \cup B| = |A' \cup B'|$, if the latter cardinals exist. Theorems already proved establish that there are bijections f from A onto A' and g from B onto B': it is then straightforward to see that $f \cup g$ is a bijection from $A \cup B$ to $A' \cup B'$, whence if either of the cardinals $|A \cup B|$ and $|A' \cup B'|$ are defined as natural numbers, the other must also be defined and must be the same. So the definition succeeds, in the sense that the value determined for $m \oplus n$ (if any) will not depend on the choice of the sets A and B. Notice that the concrete definition above is thus equivalent: for any A, B, we have $A \sim A \times \{0\}$, $B \sim B \times \{1\}$, and $A \times \{0\} \cap B \times \{1\}$ empty. That $m \oplus n$ is always defined follows from the next result:

Theorem: For any natural numbers m, n, $m \oplus n = m + n$.

Proof: We prove this by induction on n. If $n = 0$, we have $|A| \oplus 0 = |(A \times \{0\} \cup \emptyset \times \{1\}| = |A \times \{0\}| = |A|$ for any finite set A, and of course $|A| + 0$ is also $|A|$.

Suppose that the result is true for $n = k$. Let $|B| = k + 1$, so $B = C \cup \{x\}$ with $|C| = k$ and $x \notin C$. For any finite set A we have $|A| \oplus (k+1) = |(A \times \{0\}) \cup (B \times \{1\}| = |(A \times \{0\} \cup C \times \{1\}) \cup \{(x, 1)\}| = |A \times \{0\} \cup C \times \{1\}| + 1 = (|A| \oplus |C|) + 1 = (|A| \oplus k) + 1) = (|A| + k) + 1) = |A| + (k + 1)$ as desired.

Theorem: For any natural numbers m, n, $m \otimes n = m \cdot n$.

Proof: This is left as an exercise.

While we are in the spirit of elementary school, we do frame an

Abstract definition of multiplication: For any m, n natural numbers, define $m \otimes n$ as $|\bigcup X|$, where X is a pairwise disjoint collection with $|X| = n$ and $(\forall A \in X : |A| = m)$. (This could also be stated for general sets if we had a general definition of cardinality, though only if the Axiom of Choice is assumed, as we will discuss later). The alert reader

may recall that we ran into serious trouble trying to frame this definition in the type theory of chapter 2, but here it seems more innocent. However, where sets of sets are being invoked, presuming innocence may not be the best strategy. We recommend using the concrete definition above to prove the previous theorem.

We note that we get much nicer proofs (basically elementary school proofs) of a number of well known properties of arithmetic. Each of the parts of the following theorem can be proved by unpacking arithmetic operations using the $\oplus$ and $\otimes$ definitions, then explicitly constructing bijections which witness the stated equations.

Theorem: The following "axiomatic" assertions of arithmetic hold. We feel free to use the usual symbols for addition and multiplication.

commutative laws: $|A| + |B| = |B| + |A|$; $|A| \cdot |B|$

associative laws: $(|A|+|B|)+|C| = |A|+(|B|+|C|)$; $(|A|\cdot|B|)\cdot|C| = |A| \cdot (|B| \cdot |C|)$

distributive law: $|A| \cdot (|B| + |C|) = |A| \cdot |B| + |A| \cdot |C|$

identity properties and zero law: $|A|+0 = |A|$; $|A|\cdot 0 = 0$; $|A|\cdot 1 = |A|$

sample proof: We prove the distributive law. $|A| \cdot (|B| + |C|)$ is the cardinality of the collection of pairs (a, d) where $d \in B \times \{0\} \cup C \times \{1\}$, that is the collection of pairs of one of the forms $\langle a, \langle b, 0\rangle\rangle$ or $\langle a, \langle c, 1\rangle\rangle$ with $a \in A, b \in B, c \in C$. On th eother hand, $|A| \cdot |B| + |A| \cdot |C|$ consists of elements of one of the forms $\langle\langle a, b\rangle, 0\rangle$ or $\langle\langle a, c\rangle, 1\rangle$. The bijection witnessing the claimed equation of cardinals is then

$$\{\langle\langle a, \langle d, i\rangle\rangle, \langle\langle a, d\rangle, i\rangle\rangle \mid a \in A \wedge ((d \in B \wedge i = 0) \vee (d \in C \wedge i = 1))\}.$$

It is important to note that our definitions of addition and multiplication of not necessarily finite cardinals will be the same, and the proof of the previous theorem will carry over to possibly infinite cardinals. Not all sensible results of arithmetic do carry over, however. Notably, the cancellation property of addition does not hold. This can be seen informally by considering that $|1 + \mathbb{N}| = |0 + \mathbb{N}|$ will be expected to be true, as a bijection from

$\{(1,0)\} \cup \mathbb{N} \times \{1\}$ to $(\emptyset \times \{0\}) \cup (\mathbb{N} \times \{1\}) = \mathbb{N} \times \{1\}$ is readily constructed (map $\langle 1,0\rangle$ to $\langle 0,1\rangle$ and map each $\langle n,1\rangle$ to $\langle n+1,1\rangle$), but of course $1 \neq 0$. However, the following restricted form of cancellation does hold for general cardinals:

Theorem: For any sets A, B, $|A| + 1 = |B| + 1$ implies $|A| = |B|$.

Proof: We are given a bijection f from a set $A \cup \{x\}$ (with $x \notin A$) to a set $B \cup \{y\}$ (with $y \notin B$). If $f(x) = y$, the restriction of f to A witnesses $|A| = |B|$. If $f(x) \neq y$, the map $f \setminus \{\langle x, f(x)\rangle\} \setminus \{\langle f^{-1}(y), y\rangle\} \cup \{f^{-1}(y), f(x)\}$ is a bijection witnessing $|A| = |B|$. Note that this depends in no way on A and B being finite sets.

Theorem: For any sets A, B and $n \in \mathbb{N}$, $|A| + n = |B| + n \rightarrow |A| = |B|$.

Proof: Apply the previous result repeatedly.

3.3.3 The positive rationals

We now begin the construction of the real number system. It is usual to begin with the system $\mathbb{N}$ of natural numbers (which we have in hand), then construct the system $\mathbb{Z}$ oof integers, then construct the system $\mathbb{Q}$ of rationals, then construct the reals from the rationals.

We will take a somewhat different approach, perhaps reminiscent of ancient Greek prejudices against zero and negative numbers. We will begin by restricting our attention to $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$, the system of positive integers. We will then construct the system of *fractions* representing the positive rational numbers $\mathbb{Q}^+$. We will then construct the system of *magnitudes* representing the positive real numbers $\mathbb{R}^+$, and only at the final step will we introduce zero and the negative reals to obtain $\mathbb{R}$. We will see advantages in the simplicity of the definitions of arithmetic operations.

Further, we will not make any use of equivalence classes as is usual in these constructions, because choosing representative elements is easy at each point where an equivalence class construction would seem to be recommended.

Definition: For each $m, n \in \mathbb{N}^+$, we define the fraction $\frac{m}{n}$ as

$$\langle m \,\mathtt{div}\, \mathtt{gcd}(m,n), n \,\mathtt{div}\, \mathtt{gcd}(m,n) \rangle \,.$$

The development of the theory of greatest common denominators (and of the operator `div` of integer division) in the positive natural numbers should present no difficulties for our reader. We define $\mathbb{Q}^+$ as the set of all fractions. We will as usual in this sort of construction regard $\langle n, 1 \rangle$ as the fraction implementing the positive natural number n, though it is not the same object.

Definition: We define $\frac{m}{n} + \frac{p}{q}$ as $\frac{mq+np}{nq}$. We define $\frac{m}{n} \cdot \frac{p}{q}$ as $\frac{mp}{nq}$. We define $\frac{m}{n} \leq \frac{p}{q}$ as $mq \leq np$. There are things to verify here: one needs to check that choice of m, n, p, q in the representations of fractions does not affect these definitions.

The reader should find it easy to believe that we have implemented the familiar system of positive rational numbers in our set theory at this point. To check this in detail is an exercise in algebra, not set theory.

3.3.4 Magnitudes (positive reals)

This is the center of the construction, where we employ Dedekind's device of "cuts" for representing reals as sets of rationals. We only use the left set of the Dedekind cut, and we derive substantial formal advantages from only choosing to construct the positive reals in this way.

Definition: We say that a set $r \subseteq \mathbb{Q}^+$ is a *magnitude* if and only if (1) r is *nontrivial* (r and $\mathbb{Q}^+ - r$ are both nonempty), (2) r is *downward closed* (for each $p \in r$ and $q \leq p$ we also have $q \in r$), and (3) r is *open*: r has no largest element. We define $\mathbb{R}^+$ as the set of all magnitudes. For each rational p, $\{q \in \mathbb{Q}^+ \mid q \leq p \wedge q \neq p\}$ is taken to be the magnitude implementing the positive rational p, though it is not the same object. In general, speaking as if we had informal prior knowledge of the real numbers, the trick is that we implement a positive real r as $\{q \in \mathbb{Q}^+ \mid q < r\}$.

Definition: For magnitudes r, s, we define $r + s = \{p + q \mid p \in r \wedge q \in s\}$, $r \cdot s = \{p \cdot q \mid p \in r \wedge q \in s\}$, and $r \leq s$ as $r \subseteq s$.

It is valuable to note that for any nonempty set A of magnitudes which is bounded above, $\sup A = \bigcup A$, and for any nonempty set of magnitudes A which is bounded below, $\inf A = \bigcap A$. Actually, the statement about set intersections of sets of reals and greatest lower bounds is not quite true: this is the subject of an exercise.

The reader should find it easy enough to believe that we have implemented the positive reals at this point: verifying this would represent quite a lot of work in the general area of analysis.

3.3.5 The real number system

The final step of implementation of the real number system amounts to allow free rein to the operation of subtraction, and it should be reminiscent of the construction of the fractions.

Definition: For each pair of magnitudes m, n, the formal difference $m - n$ is defined as $\langle (1 + m) \ominus \mathtt{min}(m, n), (1 + n) \ominus \mathtt{min}(m, n) \rangle$, where the partial operation $\ominus$ of subtraction of magnitudes is defined in the natural

way. The set $\mathbb{R}$ of all real numbers is defined as the set of all formal differences of magnitudes. We implement 0 (this is not of course the familiar natural number 0) as $\langle 1, 1\rangle$, $+m$ as $\langle 1+m, 1\rangle$ and $-m$ as $\langle 1, 1+m\rangle$ (in which 1 is the magnitude 1, not the familiar natural number).

Definition: We define $(m-n)+(p-q)$ as $(m+p)-(n+q)$. We define $(m-n)\cdot(p-q)$ as $(mp+nq)-(mq+np)$. We define $(m-n) \leq (p-q)$ as $m+q \leq n+p$.

That the real number system has been implemented successfully at this point is an algebraic exercise.

The notations $\mathbb{N}^+, \mathbb{Q}^+$ and $\mathbb{R}^+$ might sometimes be used to describe subsets of the real number system with which their elements are "identified". The more usual systems $\mathbb{Z}$, $\mathbb{Q}$ must be understood as subsets of $\mathbb{R}$ in our development, as they are not way stations on the road to constructing the reals in our particular approach.

3.4 Preliminaries for transfinite arithmetic of cardinals and ordinals

We will now depart from implementation of familiar bits of mathematics and strike out into the uncharted territory of the infinite.

We intend to generalize the notation of cardinal $|A|$ which we have defined for all finite sets A to all sets. We state our

Formal Intention: With each set A we intend to associate a set $|A|$, the cardinality of A, satisfying the condition that for all sets A, B we have $A \sim B \leftrightarrow |A| = |B|$.

We intend to define $|A| + |B|$ as $|(A \times \{0\}) \cup (B \times \{1\})|$ and $|A| \cdot |B|$ as $|A \times B|$ for all sets A, B, as signalled above.

It happens that this intention cannot be realized in Zermelo set theory without additional assumptions. It is however possible to define $[X]^{|A|}$ as $\{B \in \mathcal{P}(X) \mid A \sim B\}$, which gives a representation of cardinals for subsets of any fixed set X.

We do immediately define the cardinality $|\mathbb{N}|$ as the set $\mathbb{N}$ itself, though when this set is considered as a cardinal we will write it $\aleph_0$ (which is read "aleph-null"). Sets of cardinality $\aleph_0$ are called *countably infinite sets*. Sets which are neither finite nor countably infinite will be called *uncountable* or *uncountably infinite* sets.

The details of implementations of cardinality will be given below.

We begin by discussing some facts about the cardinal $\aleph_0$.

Theorem: $\aleph_0 + \aleph_0 = \aleph_0$; $\aleph_0 \cdot \aleph_0 = \aleph_0$.

Proof: To prove each of these statements, we need to exhibit an appropriate bijection.

A bijection from $\mathbb{N}$ to $\mathbb{N} \times \{0\} \cup \mathbb{N} \times \{1\}$ is defined by $f(2n) = \langle n, 0\rangle$; $f(2n+1) = \langle n, 1\rangle$. This justifies $\aleph_0 + \aleph_0 = \aleph_0$.

A bijection from $\mathbb{N}$ to $\mathbb{N} \times \mathbb{N}$ is defined thus: $g(2^m(2n+1)-1) = \langle m, n\rangle$. Each positive integer factors uniquely into a product of a power of two and an odd number from which "coordinates" can be extracted as indicated; subtracting one covers all the natural numbers. This justifies $\aleph_0 \cdot \aleph_0 = \aleph_0$.

This much was known in the Middle Ages or even in ancient times. However, the former understanding was that infinite sets behaved in paradoxical ways, but all infinite sets were infinite in the same sense. It was a modern discovery that there are actually different sizes of infinite sets.

Definition: We define comparison relations between cardinals. We say that $A \preceq B$ holds iff there is an injection from A to B, and that $A \preceq^* B$ holds iff A is empty or there is a surjection from B onto A; these are two different ways of saying that the set A is no larger than the set B. We then intend to assert $|A| \leq |B|$ iff $A \preceq B$, and $|A| \leq^* |B|$ iff $A \preceq^* B$. We define $A \prec B$ as holding if $A \preceq B \wedge A \not\sim B$.

Important observations need to be made whose details will be filled in later. The starred forms are equivalent to the unstarred forms in the presence of the Axiom of Choice. The unstarred forms are preferable: there is a nice theorem (not requiring Choice) to the effect that $|A| \leq |B| \wedge |B| \leq |A| \rightarrow |A| = |B|$ (the Schröder-Bernstein theorem). In the absence of Choice, the starred version has no such nice property. The relation $\leq$ on cardinals is clearly reflexive and transitive; the theorem establishes that it is at least a partial order. We will see later that the assertion that the the order on cardinals is a total order is equivalent to the Axiom of Choice.

We define $|A| < |B|$ as $|A| \leq |B| \wedge |A| \neq |B|$.

Schröder-Bernstein Theorem: If $A \preceq B$ and $B \preceq A$, then $A \sim B$. Thus $|A| \leq |B|$ and $|B| \leq |A|$ together imply $|A| = |B|$.

Proof: Let $f : A \rightarrow B$ and $g : B \rightarrow A$ be injective. The map f sends A to a subset $f\text{“}A$ of B, the same size as A. The idea is to adjust f so that it maps A exactly to B. Let C be the set $\{(f \circ g)^n(c) \mid c \in B \setminus f\text{“}A \wedge n \in \mathbb{N}\}$. Define h as $(f \circ g)^{-1}$ on C and as the identity on $B \setminus C$. Then $h \circ f$ is a bijection from A onto B.

Now we will demonstrate the following result which was a historic surprise:

Theorem: For every set A, $|A| < |\mathcal{P}(A)|$.

Proof: Clearly $|A| \leq |\mathcal{P}|(A)|$: the map sending $a \in A$ to $\{a\} \in \mathcal{P}(A)$ witnesses this.

Suppose that $|A| = |\mathcal{P}(A)|$. This would give us a bijection $f : A \to \mathcal{P}(A)$. Now define the set $R = \{a \in A \mid a \notin f(a)\}$. Let $r = f^{-1}(R)$. Now observe that $r \in R$ iff $r \notin f(r) = R$, a contradiction. Notice the close relationship to Russell's paradox here.

So we have established $|A| < |\mathcal{P}(A)|$.

A particular consequence of this is $|\mathbb{N}| < |\mathbb{R}|$. The key is that it is straightforward to establish $\mathbb{R} \preceq \mathcal{P}(\mathbb{N})$ and $\mathcal{P}(\mathbb{N}) \preceq \mathbb{R}$ [we leave construction of the required injections as an exercise: think about representations of the reals in base 2 (though base 3 might be more convenient to avoid unwanted identifications)], so these two sets have the same cardinality, and the theorem above shows $|\mathbb{N}| < |\mathcal{P}(\mathbb{N})|$.

This is as far as we'll go with the theory of cardinals for now. As it happens the usual official representation of cardinals depends on the prior development of a representation for ordinals, and further relies on the Axiom of Choice. So we will start discussing well-orderings and ordinal numbers.

Definition: A *well-ordering* is a binary relation $\leq$ which is reflexive, anti-symmetric, transitive and total (a linear order), and has the further property that for each nonempty subset A of $\texttt{fld}(\leq)$ there is an element a such that $(\forall b \in A : a \leq b)$, a $\leq$-minimal element of A. Clearly this element is unique.

Familiar well-orderings: Of the orders encountered in undergraduate mathematics before this point, remarkably few are well-orderings. A linear order on a finite set is a well-ordering. It is amusing to observe that a relation $\leq$ and its inverse $\leq^{-1}$ are both well-orderings iff $\leq$ is a linear order on a finite set.

The usual order on the natural numbers is a well-ordering: suppose that $A \subseteq \mathbb{N}$ has no minimal element; clearly $0 \notin A$, as if otherwise 0 would be minimal in A (and for all $m \leq 0$, $m \notin A$, though it may seem odd for us to say this); now suppose for a fixed k not only that $k \notin A$, but that for all $m \leq k$, $m \notin A$: it follows that $k + 1 \notin A$ because otherwise it would be minimal in A, and further $m \notin A$ for any $m \leq k + 1$. We have proved by induction that a subset of $\mathbb{N}$ with no minimal element in the usual order is empty, and so this order is a well-ordering.

A final sort of order familiar to undergraduates which is a well-ordering is the usual order on a convergent increasing sequence of real numbers together with its limit.

The reader should check to their own satisfaction that the usual orders on integers, rationals and reals are not well-orderings.

Isomorphism: If R and S are binary relations, we say that R and S are *isomorphic* (written $R \approx S$) iff there is a bijection f from $\mathtt{fld}(R)$ onto $\mathtt{fld}(S)$ with the property that for all $x, y \in \mathtt{fld}(R)$, $x\,R\,y \leftrightarrow f(x)\,S\,f(y)$. The notion of isomorphism captures the idea that the relations R and S have the same formal structure.

We now state the formal intention which will lead to our eventual definition of ordinal numbers.

Formal Intention: We intend to associate with each well-ordering $\leq$ a set $\mathtt{ot}(\leq)$, called the *order type* of $\leq$, with the rule that for any well-orderings $\leq_1$ and $\leq_2$ we have $\mathtt{ot}(\leq_1) = \mathtt{ot}(\leq_2)$ iff $[\leq_1] \approx [\leq_2]$. Note the similarity to our formal intentions with regard to cardinality. A set which is an order type will also be called an *ordinal number*.

Reminder of segment notations: When $\leq$ represents a well-ordering, we will let $x < y$ represent $x \leq y \wedge x \neq y$. When $\leq$ is a well-ordering and $x \in \mathtt{fld}(\leq)$, we define $\mathtt{seg}_{\leq}(x)$ as $\{y \in \mathtt{fld}(\leq) \mid y < x\}$. We define $(\leq)_x$ as $[\leq] \cap \mathtt{seg}_{\leq}(x)^2$. Note that $(\leq)_x$ is also a well-ordering.

Foreshadowing of the ordinal definition: We will indicate the usual definition of order type (due to von Neumann). The difficulty is that it cannot be proved in Zermelo set theory that all well-orderings have order types in this sense:

$$\mathtt{ot}(\leq) = \{\mathtt{ot}((\leq)_x) \mid x \in \mathtt{fld}(\leq)\}$$

.

The reader should be able to convince themselves that for each natural number n, the order type of a linear order on a set of size n under this definition is precisely the von Neumann natural number n (the set of all smaller natural numbers). Further, the order type of the usual order on the natural numbers is $\mathbb{N}$ itself: we write ω for $\mathbb{N}$ when we consider

it the first infinite ordinal number. The order type of the convergent series with limit is $\{0, 1, 2, \ldots, \omega\}$, which we will call $\omega + 1$.

The embarrassment in Zermelo set theory is that the ordinal $\omega \cdot 2$ which is the order type of the order on natural numbers defined by "$m \leq_{\texttt{bogus}} n$ iff m is even and n is odd, or m and n have the same parity and $m \leq n$" cannot be shown to exist. It is easy enough to describe this set: we first define $\alpha + 1$ for any ordinal as $\alpha \cup \{\alpha\}$: this is the same as the successor definition for von Neumann naturals, and it is clear that if α is the order type of an particular order, $\alpha + 1$ will be the order type of an order extending that order by adding one more item at the end. Now $\omega \cdot 2$ is nothing more mysterious than

$$\{0, 1, 2, 3, \ldots, n, \ldots, \omega, \omega + 1, \omega + 2, \omega + 3, \ldots, \omega + n \ldots\}$$

We demonstrate briefly that $\omega \cdot 2$ cannot be shown to exist. The idea is that we can suppose that the universe consists exactly of the elements of the sets $\mathbb{N}$, $\mathcal{P}(\mathbb{N})$, $\mathcal{P}^2(\mathbb{N}), \ldots$, and no other sets: it is fairly straightforward to determine that the axioms of Zermelo set theory hold in this structure. Then observe that the ordinal ω appears first in $\mathcal{P}(\mathbb{N})$ and more generally the ordinal $\omega + i$ appears first in $\mathcal{P}^{i+1}(\mathbb{N})$. There is no iterated power set of the set of natural numbers which contains all the elements of the von Neumann ordinal $\omega \cdot 2$, so this von Neumann ordinal does not appear in this structure.

We can say this more precisely.

Iterated power sets of $\mathbb{N}$: We define $x = \mathcal{P}^i(\mathbb{N})$ as meaning "There is a sequence s with domain $i + 1$ such that $s(0) = \mathbb{N}$, for each $j < i$ we have $s(j + 1) = \mathcal{P}(s(j))$, and $x = s(i)$".

***Axiom of restriction:** For every set x, there is a natural number i such that $x \in \mathcal{P}^i(\mathbb{N})$.

Observations: We will not assume the axiom of restriction. But it is closely related to the idea of the cumulative hierarchy which we will soon introduce. Note that if $\mathbb{N}$ is defined as the Zermelo natural numbers, it follows from the axiom of restriction that the set N of von Neumann natural numbers does not exist (each von Neumann natural exists, but they do not all belong to any fixed iterated power set of the natural

numbers). Similarly, if $\mathbb{N}$ is defined as the von Neumann natural numbers, our official position, the axiom of restriction implies that the set of Zermelo natural numbers does not exist.

Further, the axiom of restriction allows a definition of cardinality! Define $|A|$ as $\{B \in \mathcal{P}^{\mathtt{sup}(\{i+1:A \not\preceq \mathcal{P}^i(\mathbb{N})\})}(\mathbb{N}) \mid A \sim B\}$. To read this, recognize that $\mathtt{sup}(\{i+1 : A \not\preceq \mathcal{P}^i(\mathbb{N})\})$ is the successor of the smallest natural number j such that $A \preceq \mathcal{P}^j(\mathbb{N})$. We define $|A|$ as the set of all sets B which are equinumerous with A and which belong to the first iterated power set of $\mathbb{N}$ which contains a set that large. Similarly we could define $\mathtt{ot}(\leq)$ as $\{B \in \mathcal{P}^{\mathtt{sup}(\{i+4:\mathtt{fld}(\leq) \not\preceq \mathcal{P}^i(\mathbb{N})\})}(\mathbb{N}) \mid [\leq] \approx B\}$: define $\mathtt{ot}(\leq)$ as the set of all binary relations isomorphic to $\leq$ on the first iterated power set of the natural numbers which is large enough to support such a relation. Under this definition, there is no difficulty defining $\omega \cdot 2$ (or much larger ordinal numbers): it is the collection of all well-orderings on subsets of ω which are isomorphic to $<_{\mathtt{bogus}}$, which appears in a fairly low indexed iterated power set of the natural numbers. This paragraph indicates that the problem with representing ordinals and cardinals in Zermelo is not that it is not strong enough; the axiom of restriction does not make Zermelo set theory stronger; it is more that the axioms of Zermelo set theory are not precise enough about the structure of the world of sets. The axiom of restriction does make Zermelo set theory more precise in this respect. The axioms we will introduce later both make the picture of the world of sets more precise and make the theory considerably stronger, in the sense that the extended theory proves the existence of more and bigger sets.

At this point our aim is to prove some facts about the structure of well-orderings.

Theorem: Any downward closed subset of the field of a well-ordering $\leq$ is a segment $\mathtt{seg}_{\leq}(x)$ or the whole of $\mathtt{fld}(\leq)$.

Theorem: No well-ordering $\leq$ is isomorphic to any of its segment restrictions $(\leq)_x$.

Theorem: For any well-orderings $\leq_1$ and $\leq_2$ exactly one of the following is true:

1. $\leq_1 \approx \leq_2$.
2. For some $x \in \mathtt{fld}(\leq_1)$, $(\leq_1)_x \approx \leq_2$.
3. For some $x \in \mathtt{fld}(\leq_2)$, $\leq_1 \approx (\leq_2)_x$.

Proofs: The proofs of all of these statements are *exactly* as in section 2.12.

Definition: For ordinals α and β, we define $\alpha \leq_\Omega \beta$ as holding iff for some (and so for any) well-orderings $\leq_1, \leq_2$ such that $\alpha = \mathtt{ot}(\leq_1)$ and $\beta = \mathtt{ot}(\leq_2)$ we have $\leq_1$ isomorphic either to $\leq_2$ or to some segment restriction $(\leq_2)_x$ for $x \in \mathtt{fld}(\leq_2)$. We will usually write just $\leq$ for $\leq_\Omega$ except where some confusion might otherwise occur.

Theorem: The relation $\leq_\Omega$ is reflexive, antisymmetric, transitive, and total, and for any set of ordinals A, there is a $\leq_\Omega$-minimal ordinal in A. We do not say that $\leq_\Omega$ is a well-ordering only because it is not a set, as we will see shortly.

Proof: This is proved in section 2.12, and the proof can be read using our definitions and works in exactly the same way.

Theorem: The order type of the restriction of $\leq_\Omega$ to the set of ordinals β such that $\beta < \alpha$ is α, if this set exists.

Proof: Assume that $\{\beta \mid \beta <_\Omega \alpha\}$ exists. From this it follows that $[\leq_\Omega] \cap \{\beta \mid \beta <_\Omega \alpha\}^2$ exists. Suppose that this statement were not true for some α. Then we can argue that there is a smallest ordinal γ for which it is not true. If $\{\beta \mid \beta < \alpha\}$ contains an ordinal for which the theorem is not true, then the smallest element γ of $\{\beta \mid \beta < \alpha\}$ is the smallest counterexample. Otherwise $\gamma = \alpha$ itself is the smallest counterexample. So, let γ be the smallest ordinal such that the order type of the restriction of $\leq_\Omega$ to ordinals less than γ is not γ. Let γ' be the actual order type of this restriction. For each $\delta < \gamma$, the order type of the restriction of $\leq_\Omega$ to ordinals $< \delta$ has order type δ, and so we see that every $\delta < \gamma$ is also $< \gamma'$, from which it follows that $\gamma \leq \gamma'$. If $\gamma < \gamma'$, it follows that some segment restriction of the natural well ordering on ordinals $< \gamma$ must be of order type γ, but this

is impossible: each such segment restriction is the order type of the natural well-ordering on ordinals $<$ some $\delta < \gamma$, and this order type is by choice of γ equal to $\delta < \gamma$. So the order type of the natural order on the ordinals $< \gamma$ is γ, which is a contradiction.

† It is important to note that the statement proved here cannot be proved in type theory and in fact would not make sense in type theory, because the ordinal α would occur at more than one type.

Theorem: There is no set of all ordinals, and $\leq_\Omega$ is not a set.

Proof: If there were a set $\mathtt{Ord}$ of all ordinals, then $\leq_\Omega$ would be a set, and in fact a well-ordering, and so there would be an ordinal $\Omega = \mathtt{ot}(\leq_\Omega)$. We would have $\Omega \in \mathtt{Ord} = \mathtt{fld}(<_\Omega)$. By the previous theorem, for any ordinal α, we have $\mathtt{ot}((\leq_\Omega)_\alpha) = \alpha$ so in particular $\mathtt{ot}((\leq_\Omega)_\Omega) = \Omega$. But equally clearly $\mathtt{ot}((\leq_\Omega)_\Omega)$ is strictly less than $\mathtt{ot}(<_\Omega) = \Omega$; a well-ordering cannot be isomorphic to one of its segment restrictions.

We now develop our official definition of ordinals and state an axiom required to make it work.

Definition: A *system of von Neumann ordinal notation* is a function f whose domain is the field of a well-ordering $\leq$ and which satisfies the condition $f(x) = \{f(y) \mid y \leq x \wedge y \neq x\}$ for all x in the domain of $\leq$.

Theorem: If f is a system of von Neumann ordinal notation on the field of $\leq$ and $f(x) = f(y)$, then $x = y$: systems of von Neumann ordinal notation are injective.

Proof: Let x be the $\leq$-minimal element of the field of $\leq$ such that $f(x) = f(y)$ for some $y \neq x$. Clearly $x \leq y$. We are supposed to have

$$\{f(z) \mid z \leq x \wedge z \neq x\} = \{f(w) \mid w \leq y \wedge w \neq y\}.$$

Now observe that $f(x)$ belongs to the second set but cannot belong to the first.

Theorem: If well-orderings $\leq$ and $\leq'$ are isomorphic and support systems of von Neumann ordinal notation f and g, then for each x in the domain of $\leq$, $f(x) = g(y)$ iff $(\leq)_x \approx (\leq)_y$.

Proof: Consider the smallest x for which this is not the case. We have $(\leq)_x$ isomorphic to $(\leq)_{h(x)}$, this being witnessed by the restriction of h to the segment determined by x. We have $f(z) = g(h(z))$ for each $z \leq x$. And it follows that $f(x) = \{f(z) \mid z \leq x\} = \{g(h(z) \mid z \leq x\} = \{g(w) \mid w \leq' h(z)\} = g(h(x))$. On the other hand, if $f(x) = g(y)$ for some $y \in \mathtt{fld}(\leq')$, we have $f(x) = \{f(z) \mid z \leq x\} = \{g(h(z)) \mid z \leq x\} = g(h(z))$, whence $f(x) = g(y)$ since systems of von Neumann ordinal notation are injective.

Axiom of Ordinals: On every well-ordering $\leq$, there is a system of von Neumann ordinal notation.

Remark about the Axiom of Ordinals: This axiom or something stronger is needed: the ordinal $\omega \cdot 2 + 1$ can be shown not to support a system of von Neumann ordinal notation under the Axiom of Restriction. The Axiom of Ordinals will be seen to be a consequence of the more powerful Axiom of Replacement that we will adopt later.

Definition: We define $\mathtt{ot}((\leq)_x)$ as $f(x)$, where f is a system of von Neumann ordinal notation on $\leq$. It is straightforward to verify that every well-ordering $\leq$ can be expressed in the form $(\leq')_x$ for some "larger" well-ordering $\leq'$, and straightforward to establish that the value of $\mathtt{ot}(\leq)$ computed by the definition above does not depend on the choice of $\leq'$. We refer to order types defined in this way as *von Neumann order types*, and we refer to any von Neumann order type as a *von Neumann ordinal number* or (usually) simply as an ordinal number.

Exercises

1. Using the official definititions of $[X]^n$ and $[X]^{<\omega}$, prove that the von Neumann natural number n is an n element set (that $n \in [\mathbb{N}]^n$ is probably easiest to prove). Perhaps much harder, prove that $\mathbb{N}$ is an infinite set, using the official definition of $[X]^{<\omega}$. This all hinges on details of definitions, and in no way on common sense!

2. Prove the theorem $m \otimes n = m \cdot n$, verifying that the set based definition of multiplication is equivalent to the Peano arithmetic definition. You may use the result about addition already established in the notes in your proof.

3. Present the proofs of the associativity of addition and multiplication of cardinal numbers of sets in the same style in which I presented the proof of the distributive property of multiplication over addition. An adequate description of the bijection witnessing the claimed equation between cardinals in each case is all that is wanted.

4. Prove that the sum of two magnitudes is a magnitude. You may assume all familiar properties of positive rational numbers (the elements of magnitudes). You have a definition of the sum of two magnitudes: the point is to show that this set satisfies the defining conditions to be a magnitude.

5. I claimed (correctly) in class and in the notes that the supremum of a nonempty set A of magnitudes which is bounded above is the union $\bigcup A$ of the set A. I also claimed that the infimum of a nonempty set A of magnitudes which is bounded below is the intersection $\bigcap A$ of the set A. This statement about the infimum (greatest lower bound) is not quite true! Describe an exception and indicate how to correct this statement. This hinges on details of the definition of magnitude.

6. Identify the smallest natural number n such that $\mathbb{R} \in \mathcal{P}^n(\mathbb{N})$. This should be easy bookkeeping.

7. Describe an injection from $\mathbb{R}$ into $\mathcal{P}(\mathbb{N})$ and an injection from $\mathcal{P}(\mathbb{N})$ into $\mathcal{R}$. These maps do not need to be onto: the point is to show that each set is the same size as a subset of the other. Ordinary knowledge of the reals is all that is needed (this doesn't depend in any way on my fancy constructions). Think about base 2 (or, for reasons having to do with unintended identifications, base 3) representations of the reals.

8. Prove that a linear order $\leq$ has finite domain iff both $\leq$ and $\leq^{-1}$ are well-orderings.

3.5 Zorn's Lemma, The Well-Ordering Theorem, and the official definition of cardinality

In this section we will realize Zermelo's aim in the definition of his axioms for set theory: we will prove the Well-Ordering Theorem, that every set is the field of a well-ordering. Amachronistically, we will do this by proving Zorn's Lemma, a theorem whose proof is formally quite similar to the Well-Ordering Theorem, from which the Well-Ordering Theorem is easily proved, and which is technically very useful in set theory and in mathematics generally.

Definition: Fix a partial order $\leq$. A *chain* in c is defined as a linear order which is a subset of $\leq$. An upper bound for a chain $\leq_c$ in $\leq$ is an $x \in \mathtt{fld}(\leq)$ such that for all $c \in \mathtt{fld}(\leq_c)$ we have $c \leq x$. A *maximal element* for $\leq$ is an $m \in \mathtt{fld}(\leq)$ such that for all n, $m \leq n$ implies $m = n$.

Definition (ordinal indexing): Let $\leq$ be a well-ordering and let α be an ordinal. We introduce the notation $[\leq]_\alpha$ for the unique x, if there is one, such that the order type of $(\leq)_x$ is α.

Zorn's Lemma: Let $\leq$ be a partial order with the property that every chain in $\leq$ has an upper bound. Then $\leq$ has a maximal element.

Proof of the Well-Ordering Theorem from Zorn's Lemma: Let A be a set. Let X be the set of well-orderings whose fields are subsets of A. Define a partial order on X by $[\leq_1] \leq [\leq_2]$ iff $[\leq_1] = [\leq_2] \vee (\exists a \in A : [\leq_1] = [(\leq_2)_a])$: that is, if $\leq_2$ is an end extension of $\leq_1$. Every chain in this order on X has a well-ordering of a subset of A as the union of its field, which is an end-extension of each element of the union of its field. So there is a maximal element in the end extension order on X by Zorn's Lemma, and this maximal element must be a well-ordering of all of A: any order on a subset of A whose field is not all of A can be end-extended by adding another element of A as a new largest element.

Proof of Zorn's Lemma: Let $\leq$ be a partial order in which every chain has an upper bound. Let X be the set of all pairs $(\leq_c, x)$, where x is an upper bound for $\leq_c$, and moreover $x \in \mathtt{fld}(\leq_c)$ if and only if

x is maximal in $\leq$ (clearly we can always choose an upper bound for $\leq_c$ which is not in the field of $\leq_c$, except in the case where $\leq_c$ has a maximal element which is also a maximal element in $\leq$). We now define Y as the partition of X whose elements are the sets

$$Y_{\leq_c} = \{\langle \leq_c, x\rangle \mid \langle \leq_c, x\rangle \in X\}.$$

Now choose a choice set F for the partition Y. Notice that F is a function which sends each chain in $\leq$ to one of its upper bounds, belonging to the field of the chain only if it is maximal in $\leq$.

Define a *special chain* in $\leq$ as a chain $\leq_s$ in $\leq$ which is a well-ordering and has the property that for each x in the field of $\leq_s$ we have $F((\leq_s)_x) = x$.

We claim that if $\leq_1$ and $\leq_2$ are special chains, either the two special chains are equal or one is a segment restriction of the other. If this is not the case, it follows that there is an α such that $[\leq_1]_\alpha \neq [\leq_2]_\alpha$ (both being defined), and there will be a smallest such α. But then for this smallest α the segment restrictions $(\leq_1)_{[\leq_1]_\alpha} = (\leq_1)_{[\leq_2]_\alpha}$, from which it follows that $[\leq_1]_\alpha = F((\leq_1)_{[\leq_1]_\alpha}) = F((\leq_1)_{[\leq_2]_\alpha}) = [\leq_2]_\alpha$, which is a contradiction.

We rephrase the previous paragraph in a way which does not use ordinal indexing: we do want it to be clear that existence of von Neumann ordinals is not required for this proof. If $\leq_1$ and $\leq_2$ are special chains which are not equal, and neither of which is a segment restriction of the other, there must be a $\leq_1$-minimal x and a $\leq_2$-minimal y such that $x \neq y$ and $(\leq_1)_x = (\leq_2)_y$, whence it follows that $x = F((\leq_1)_x) = F((\leq_2)_y) = y$, which is a contradiction.

From this it follows that the set union of all special chains is itself a special chain: this is again the union of a collection of well-orderings ordered by end extension, so it is a well-ordering as noted above, and further it clearly satisfies the special chain property. Let the union of all special chains be denoted by $\leq_S$. Now consider $F(\leq_S)$: this must be an upper bound for $<_S$: if it is not in the field of $<_S$ then $<_S$ can be extended to a longer special chain by appending $F(\leq_S)$; but this is impossible because $<_S$ is the union of all special chains. Thus $F(<_S)$ is in the field of $<_S$, from which it must be maximal in $\leq$.

Now that we know that all sets can be well-ordered, we can present the official definition of cardinality.

Definition: For any set A, we define $|A|$ as the smallest ordinal α such that there is a well-ordering $\leq$ with field A such that $\mathtt{ot}(\leq) = \alpha$.

Observation: If this definition of cardinality is used, $|A| \sim A$.

Proof of observation: $|A|$ is the order type of a well-ordering with field A, and the von Neumann order type of any well-ordering is equinumerous with the field of that well-ordering.

Now we combine observations of the last two sections to raise a classic question. We have shown that there is an uncountable set (for example $\mathcal{P}(\mathbb{N})$). It follows that there is a smallest uncountable ordinal, which we will call ω_1. Clearly ω_1 is a cardinal (for example $|\omega_1| = \omega_1$): when we consider it as a cardinal we call it $\aleph_1$.

Now we can raise a famous question (Cantor's Continuum Hypothesis):

Question: Is it the case that $|\mathcal{P}(\mathbb{N})|(= |\mathbb{R}|) = \aleph_1$?

Exercises

1. Prove that the union of a collection of well-orderings which are linearly ordered by end-extension is a well-ordering. Give an example of a collection of well-orderings which are linearly ordered merely by inclusion (the subset relation) whose union is *not* a well-ordering.

3.6 The cumulative hierarchy picture and Replacement

This section outlines the development of the cumulative hierarchy picture of the world of sets, and the principle of "limitation of size" (collections are sets if they are small in a suitable sense).

3.6.1 Basic definitions of ordinal and cardinal numbers in untyped set theory; the cumulative hierarchy introduced

We now present the definitions of cardinal and ordinal number which are usually used in *ZFC*. We give those definitions (due to von Neumann) but they have the limitations that they do not necessary work in Zermelo set theory without Replacement (not all well-orderings can be shown to have order types, nor can all sets be shown to have cardinals) and the von Neumann definition of cardinal depends essentially on the Axiom of Choice, as the Scott definition does not.

Of course this section, up to the Axiom of Ordinals, redevelops the notion of von Neumann ordinal already introduced in a different style in the previous section.

The informal motivation of the von Neumann definition of natural numbers and general ordinals is the following

*__Circular Definition:__ Each ordinal is the set of all preceding ordinals.

__Development:__ Thus the first ordinal 0 is $\emptyset$, 1 is $\{0\}$, 2 is $\{0,1\}$, 3 is $\{0,1,2\},\ldots$. And further, ω is the set of all finite ordinals $\{0,1,2\ldots\}$, $\omega+1$ is $\{0,1,2,\ldots,\omega\}$, $\omega\cdot 2$ is $\{0,1,2,\ldots,\omega,\omega+1,\omega+2\ldots\}$, and so forth.

We give a formal definition with the same effect.

__Definition:__ A *transitive set* is a set A such that $A\subseteq\mathcal{P}(A)$. Equivalently, for all x, y, if $x\in y$ and $y\in A$, then $x\in A$ (this might suggest why the word "transitive" is used).

Definition: A *(von Neumann) ordinal number* is a transitive set which is strictly well-ordered by the membership relation. Equivalently, it is a set which is transitive, not a member of itself, and well-ordered by the inclusion (subset) relation.

†**Observation (depending on section 3.7.1 below):** In our implementation of Zermelo set theory in set pictures, a von Neumann ordinal number α is implemented by the isomorphism type of strict well-orderings of type $\alpha+1$ (except for 0, which is implemented by the order type of the usual empty order). In $\mathbb{E}_{T^2(\lambda)}$, only the ordinals less than λ are implemented in this way. If $\lambda = \omega \cdot 2$, this definition is not useful: the only infinite well-orderings with order types are of the form $\omega + n$, but there are much longer order types that are realized (such as ω_1). A hypothesis adequate to make this definition useful is "$\beth_{T^2(\lambda)}$ exists for each ordinal λ" in the ambient type theory. The Axiom of Replacement of *ZFC* makes this definition usable (and is much stronger).

Definition: The *(von Neumann) order type* of a well-ordering W is the von Neumann ordinal α such that the union of the restrictions of the membership and equality relations to α is isomorphic to W. Equivalently, the subset relation restricted to α is isomorphic to W.

Definition: The *(von Neumann) cardinality* of a set A is the smallest von Neumann ordinal which is the order type of a well-ordering of A.

Zermelo set theory cannot prove the existence of any von Neumann ordinals other than the finite ones, in its original formulation. In modern reformulations, the axiom of infinity is often given as asserting the existence of the von Neumann ordinal ω, in which case the first ordinal whose existence cannot be proved is $\omega \cdot 2$.

We handle this provisionally by extending Zermelo set theory with an

Axiom of Ordinals: For each well-ordering $\leq$, there is a von Neumann ordinal α (necessarily unique) which we denote by $\mathtt{ot}(\leq)$, read *the order type of* $\leq$, such that the subset relation restricted to α is isomorphic to $\leq$.

Remark: This axiom ensures that every well-ordering has a corresponding von Neumann ordinal to serve as its order type.

A further axiom gives an intuitive description of the way that we now envision the universe of sets as being built. We start with the empty set and build the universe in stages indexed by the ordinals, taking power sets at successor stages and taking unions at limit stages.

Axiom of Levels: For each ordinal α, there is a set V_α, this scheme satisfying

1. $V_0 = \emptyset$
2. $V_{\alpha+1} = \mathcal{P}(V_\alpha)$
3. for λ a limit ordinal, $V_\lambda = \bigcup\{V_\beta \mid \beta < \lambda\}$.

In addition, for every set x there is an ordinal α such that $x \in V_\alpha$.

The Axiom of Ordinals and the Axiom of Levels are not as strong as the Axioms of Replacement and Foundation usually adjoined to Zermelo set theory, but they do give a good picture of the structure which the universe of untyped set theory is usually intuitively understood as having.

We present a formulation of the construction of levels which makes it clear that the notation V_α can be defined in the language of set theory (it is not required that we introduce a new primitive notion V_α to be able to talk about the levels).

Definition: A *subhierarchy* is a set H which is well-ordered by inclusion and in which each successor in the inclusion order on H is the power set of its predecessor and each non-successor in the inclusion order on H is the union of all its predecessors in that order. A *rank* is a set which belongs to some subhierarchy.

Theorem: Of any two distinct subhierarchies, one is an initial segment of the other in the inclusion order. So all ranks are well-ordered by inclusion.

Proof: Let H_1 and H_2 be subhierarchies.

Suppose first that H_1 is included in H_2. From this it follows that H_1 is an initial segment of H_2, unless there is an element $h_2 \in H_2 \setminus H_1$ which is a subset of some $h_1 \in H_1$. Choose the minimal h_1 in the inclusion order on H_1 such that there is h_2 with this property, and choose the minimal such h_2 for the given h_1. If h_2 is a successor in the inclusion

order on H_2, then it is the power set of some $h_3 \in H_2$, and this h_3 must also belong to H_1 by minimality of h_2. But then the successor of h_3 in the inclusion order on H_1 exists (because h_1 properly includes h_2 and so h_3) and is the power set of h_3, and of course h_2 is also the power set of h_3 because it is the successor of h_3 in the order on H_2, so $h_2 \in H_1$ which is a contradiction. If h_2 is limit, it is the union of all elements of the domain of H_2 properly included in H_2, all of which actually are elements of H_1. There is a first element of H_1 properly including all of these sets, because h_1 properly includes h_2 and so properly includes all of these sets. But this first element is the union of all the elements of H_1 properly included in it, and so in fact is h_2, so $h_2 \in H_1$, again a contradiction. We have established that if H_1 is included in H_2, then H_1 is an initial segment of H_2 in the inclusion order.

Now suppose that H_1 is not included in H_2. There must then be a first h_2 in the inclusion order on H_2 such that $h_2 \notin H_1$. We claim that in this case H_1 is an initial segment of H_2 in the inclusion order. Certainly the collection H_3 of subsets of h_2 which belong to H_1 is such an initial segment. Suppose that there is $h_1 \in H_1$ which is not a subset of h_2; choose the minimal such h_1. If h_1 is a successor in the inclusion order on H_1, it must be the power set of some $h_3 \in H_1$ which is a subset of h_2. Now we argue that h_2 must also be the power set of h_3: h_3 is included in h_2, and if h_2 were not its immediate successor in H_2 then its immediate successor would be the power set of h_3 and would be included in h_2 as a subset, so h_1 would be included as a subset in h_2, which is a contradiction. But also h_1 cannot be equal to h_2 because $h_1 \in H_1$ and $h_2 \notin H_1$. We are left in the case where h_1 is the union of all elements of H_3. Now we observe that there must be a first h_4 in H_2 which includes all elements of H_3, because $h_2 \in H_2$ includes all elements of H_3, and this must be the union of all elements of H_3 as well. But this means that $h_4 \in H_2$ is equal to $h_1 \notin H_2$, which is a contradiction.[9]

[9]This is gruesome. It might make a classroom exercise? Or I may just need to rewrite it.

Alternative definition of hierarchy given in lecture: Alternatively, we define a *hierarchy along a well-ordering* $\leq$ as a function h with domain $\mathtt{fld}(\leq)$ and satisfying the following conditions:

1. The image of the $\leq$-first element of $\mathtt{fld}(\leq)$ is $\emptyset$.
2. If y is the immediate successor of x in the order $\leq$, then $h(y) = \mathcal{P}(h(x))$
3. If y is an element of $\mathtt{fld}(\leq)$ which is not an immediate successor in the order $\leq$, then $h(y) = \bigcup\{h(z) \mid z \leq y \wedge z \neq y\}$.
4. Conditions 1-3 can be replaced by the single condition "For all $x \in \mathtt{fld}(\leq)$, $h(x) = \bigcup\{\mathcal{P}(h(y)) \mid y \leq x\}$".

Comments on the alternative definition: It should be clear that the intention is that for any $x \in \mathtt{fld}(\leq)$, that $h(x) = V_{\mathtt{ot}((\leq)_x)}$. We can then give an alternative definition of rank: a *rank* as any set which belongs to the range of any hierarchy along any well-ordering. It is straightforward to prove that for any well-orderings $\leq_1$ and $\leq_2$, with hierarchies h_1 and h_2 along them, $h_1(x) = h_2(y) \leftrightarrow (\leq_1)_x \approx (\leq_2)_y$, for any x in the field of $\leq_1$ and y in the field of $\leq_2$: all hierarchies agree in a suitable sense.

Axiom of Rank: Every set is a subset of some rank.

Definition: For any formula $\phi[x]$, define $r_{\phi[x]}$ as the minimal rank r such that $(\exists x \in r.\phi[x])$, or as the empty set if there is no such rank r. Define $\{x :: \phi\}$ as $\{x \in r_{\phi[x]} \mid \phi[x]\}$. $\{x :: \phi[x]\}$ is obviously a set for all formulas $\phi[x]$.

Definition: $A \sim B$ iff there is a bijection from A onto B, as in type theory. $|A|$, the *Scott cardinal* of A, is defined as $\{B :: B \sim A\}$. For any relations R and S, we say that $R \approx S$ iff there is a bijection f from $\mathtt{fld}(R)$ onto $\mathtt{fld}(S)$ such that $x\,R\,y \leftrightarrow f(x)\,S\,f(y)$. We define the *Scott isomorphism type* of R as $\{S :: S \approx R\}$. Scott isomorphism types of well-orderings are called *Scott order types* (of the well-orderings: $\mathtt{ot}(W)$ is the Scott order type of a well-ordering W) or *Scott ordinal numbers* (as a class).

Notice that this "Scott trick" allows us to recover the ability to define isomorphism types of objects as (restricted) equivalence classes.

Now that we have defined Scott ordinals we can define the notation V_α.

Definition: For any subhierarchy h, introduce the nonce notation $\leq_h$ for the inclusion order restricted to h. If α is a Scott ordinal we define V_α as the rank A (if there is one) such that $\mathtt{ot}((\leq_h)_A) = \alpha$ for any $\leq_h$ with A in its field. It is straightforward to show that $\mathtt{ot}((\leq_h)_A)$ is the same ordinal for any $\leq_h$ with A in its field, and that A is uniquely determined by α. We can also use the alternative definition: define V_α as the value of any hierarchy along a well-ordering $\leq$ of order type $\alpha + 1$ at the $\leq$-maximum element of $\mathtt{fld}(\leq)$.

In Zermelo set theory with the Rank Axiom we can prove that every set belongs to some V_α but we cannot prove the existence of $V_{\omega \cdot 2}$. But we do have the ability to define order types for every well-ordering and cardinals for every set using the Scott definitions. And we can then recover the full Axiom of Levels as the

Axiom of Hierarchy: For each Scott ordinal α, V_α exists. Equivalently, there is a subhierarchy isomorphic to each well-ordering (or a hierarchy along each well-ordering, using the alternative definition).

It is an immediate consequence of the Axiom of Hierarchy that each von Neumann ordinal exists (the von Neumann α exists in $V_{\alpha+1}$). So the axioms of Rank and Hierarchy have the same effect as the axioms of Ordinals and Levels. Once we have the axiom of hierarchy, we can if we prefer use von Neumann ordinals instead of Scott ordinals to index V_α's and serve as order types of well-orderings.

3.6.2 More on the von Neumann definitions of ordinal and cardinal number

We introduced the perhaps mysterious traditional definition of *ordinal number* due to von Neumann above:

Definition: An *ordinal number* is a transitive set A which is strictly well-ordered by membership (i.e., the restriction

$$[\in] \cap A^2 = \{\langle x, y\rangle \in A \times A \mid x \in y\}$$

of the membership relation to A is the strict partial order corresponding to a well-ordering). Or "a transitive set of transitive sets none of which are self-membered".

Observation: This is equivalent to "x is an ordinal iff x is a transitive set, no element of x is self-membered, and x is well-ordered by inclusion". This has the merit that our preferred definition of well-ordering is used. Let x be an ordinal by this definition. For each $y \in x$, and each $z \in y$, we have $z \in x$ because x is transitive, so we have either $z \subset y$ or $y \subseteq z$. But $y \subseteq z$ is impossible because this would imply $z \in z$.

Definition: For any ordinal α, use $\in_\alpha$ to represent $[\in] \cap \alpha^2$ (which we know is a strict well-ordering).

Theorem: For any ordinal α and any $\beta \in \alpha$, β is an ordinal, $\beta = \mathtt{seg}_{\in_\alpha}(\beta)$ and $\in_\beta = (\in_\alpha)_\beta$.

Proof: $\delta \in \gamma \in \beta \rightarrow \delta \in_\alpha \gamma \in_\alpha \beta$ ($\gamma, \delta \in \alpha$ because α is transitive) and this implies $\delta \in_\alpha \beta$ and so $\delta \in \beta$ because $\in_\alpha$ is a partial order. Thus $\beta \in \alpha$ is transitive. $[\in] \cap \beta^2$ is a strict well-ordering because it is a suborder of $[\in] \cap \alpha^2$. Further, it is evident that $\in_\beta = (\in_\alpha)_\beta$: the order on β is the segment restriction of the order on α determined by β, because β is identical to the segment in the order on α determined by β: $\beta = \{\gamma \in \alpha \mid \gamma \in \beta\}$ (this uses transitivity of α) $= \{\gamma \mid \gamma \in_\alpha \beta\}$, which is what we mean by $\mathtt{seg}_{\in_\alpha}(\beta)$

Theorem: For any two ordinal numbers α and β, exactly one of the following is true: $\alpha = \beta$, $\alpha \in \beta \wedge \alpha \subseteq \beta$, $\beta \in \alpha \wedge \beta \subseteq \alpha$. Any set of ordinal numbers is thus linearly ordered by $\subseteq$: moreover, this linear order is a well-ordering.

Proof: By a basic theorem on well-orderings proved above, we know that there is either an isomorphism from $\in_\alpha$ to $\in_\beta$, an isomorphism from $\in_\alpha$ to some $(\in_\beta)_\gamma = \in_\gamma$ for some $\gamma \in \beta$ or an isomorphism from $\in_\beta$ to some $(\in_\alpha)_\gamma = \in_\gamma$ for some $\gamma \in \alpha$. It is then clearly sufficient to show that for any ordinals α and β, if $\in_\alpha \approx \in_\beta$, then $\alpha = \beta$. Suppose for the sake of a contradiction that $f : \alpha \to \beta$ is an isomorphism from $\in_\alpha$ to $\in_\beta$ and that there is some $\gamma \in \alpha$ such that $f(\gamma) \neq \gamma$. There is then a $\in_\alpha$-least such γ. We have γ as the $\in_\alpha$-least element of α such that $f(\gamma) \neq \gamma$. The objects which are $\in_\beta f(\gamma)$ are exactly those $f(\delta)$ such that $\delta \in_\alpha \gamma$ (this is just because f is an isomorphism). We can read $\in_\alpha$ and $\in_\beta$ simply as membership, and we remind ourselves that for any $\delta < \gamma$ $f(\delta) = \delta$, and thus we see that $\gamma = f(\gamma)$ because they have the same members, which is a contradiction.

That $\alpha \in \beta \to \alpha \subseteq \beta$ expresses the fact that ordinals are transitive sets. $\subseteq$ is a partial order on sets and what we have shown so far indicates that it is a linear order on ordinals. To see that it is a well-ordering, we need to show that any nonempty set A of ordinals has a $\subseteq$-least element: since A is nonempty, we can choose $\alpha \in A$; either there is some $\beta \in \alpha$ which is an element of A or there is not. If there is none, then α is the $\subseteq$- (and $\in$-) least element of A; otherwise the $\subseteq$- (and $\in$-) least element of α which belongs to A will be the $\subseteq$- (and $\in$-) least element of A: there is such an element because $\in$ is a strict well-ordering of α and so $\subseteq$ is a well-ordering of α.

Definition: For any well-ordering $\leq$, we define $\mathtt{ot}(\leq)$ as the ordinal α (if any) such that the well-ordering of α by $\subseteq$ is isomorphic to $\leq$.

Definition: For any set A, we define $|A|$, the cardinality of A, as the minimal ordinal α in the inclusion order such that $A \sim \alpha$. Ordinals which are cardinals are also called initial ordinals. For any ordinal α, we define $\beth_\alpha$ as the infinite cardinal with the property that the order type of the inclusion order on smaller infinite cardinals is α.

This definition of cardinal does not make sense unless we assume the Axiom of Choice (so that every set can in fact be well-ordered) and at least the Axiom of Ordinals (so that every well-ordering has a von Neumann order type). The Scott definition is available as an alternative if the Axiom of Rank (or the Axiom of Levels) is present. We are assuming in general in

this section that we are assuming either Ordinals and Levels, or Rank and Hierarchy (each of these pairs of axioms has the same effect).

Note that in the usual set theory we identify a cardinal number with its initial ordinal: these are not the same object in type theory, though of course they are closely related. This is another of those differences between possible implementations of mathematical concepts in set theory that one should watch out for (in the Scott implementation of cardinals and ordinals in Zermelo set theory with the Axiom of Rank, a cardinal is not identified with its initial ordinal). The fact that though we identify these concepts formally in *ZFC* we do not actually think of them as having the same mathematical content is witnessed by the fact that we use different notations for $\mathbb{N}$ (the set of natural numbers), ω (the first infinite ordinal) and $\aleph_0$ (the first infinite cardinal) although these are all implemented as exactly the same object! Note that in type theory they are all different.

It is important to notice that just as there can be no set V of all sets in Zermelo set theory, there can be no set `Ord` of all ordinals (so transfinite induction and recursion must be stated in property-based or restricted forms in this theory). For the ordinals are strictly well-ordered by membership in an obvious external sense: if there were a set Ω which contained all ordinals, it would be an ordinal, so we would have $\Omega \in \Omega$, and this is impossible again by the definition of ordinal. This is a version of the Burali-Forti paradox, another of the classical paradoxes of set theory.

Exercises

1. The Scott definition of a natural number n is that it is the collection of all sets of size n and rank as low as possible. Remember the rank of a set A is the first ordinal α such that A is a subset of V_α. Write down as many Scott natural numbers as explicit sets as you can stand to. Work out the sizes of the next few (how many elements do they have? – go up to 20 or so?) All you need for this is an understanding of what $V_0, V_1, V_2 \ldots$ (the finite ranks) are, and some familiar combinatorics. You might also want to see what you can say about the Scott natural number 60000 versus the Scott natural 70000. There is a dramatic difference (smiley).

2. The Axiom of Foundation asserts that for any nonempty set x there is a set $y \in x$ such that $x \cap y = \emptyset$.

   One way of understanding this is that this axiom says that if we look at $[\in] \cap x^2$ (the membership relation on x) that it must have a "minimal" element – "minimal" is in scare quotes because membership is not an order relation. A "minimal" element y will have empty preimage under the membership relation restricted to x – that is, it will have no elements in common with x.

   Use the Axiom of Foundation (along with the other axioms of course) to prove the following:

   (a) There is no set x such that $x \in x$.

   (b) There is no sequence s such that for all $n \in \mathbb{N}$ we have $s_{n+1} \in s_n$.

   The strategy to follow is this: in each part, identify a set which would have no "minimal" element in the membership relation.

3.6.3 The Axiom of Replacement and *ZFC*

We introduce the missing assumption of the usual set theory which makes it possible to prove that the von Neumann definitions of ordinal and cardinal number are total. The axioms of Replacement and Foundation imply the axioms of Ordinals and Levels (or Rank and Hierarchy) and are in fact considerably stronger.

class function notation: If we have a formula $\phi[x, y]$ such that for every x there is at most one y such that $\phi[x, y]$, we introduce notation $y = F_\phi(x)$ for the unique y associated with a given x. Notice that F_ϕ is not understood to be a set here.

Axiom (scheme) of Replacement: If we have a formula $\phi[x, y]$ such that for every x there is at most one y such that $\phi[x, y]$, and define $F_\phi(x)$ as above, then for every set A, the set $\{F_\phi(x) \mid x \in A\}$ exists.

This is called an axiom scheme rather than an axiom, because we actually have a distinct axiom for each formula, in a technical sense.

The Axiom of Replacement can be used then to justify the recursive definition of the V_α's above. What the axiom of replacement says, essentially, is that any collection we can show to be the same size as or smaller than a set is in fact a set.

Theorem: $\mathtt{ot}(\leq)$ exists for every well-ordering $\leq$.

Proof: Let $\leq$ be a well-ordering such that $\mathtt{ot}(\leq)$ does not exist. If there are x such that $\mathtt{ot}((\leq)_x)$ does not exist, define $\leq_0$ as $(\leq)_x$ for the smallest such x; otherwise define $\leq_0$ as $\leq$ itself. In either case $\leq_0$ is a well-ordering which has no order type with the property that all of its initial segments have order types. We now define a formula $\phi[x, \alpha]$ which says "α is the order type of $(\leq_0)_x$" (the tricky bit is showing that we can say this). Notice that once we do this we are done: $\{F_\phi(x) \mid x \in \mathtt{fld}(\leq_0)\}$ will be the first von Neumann ordinal after all the order types of segment restrictions of $\leq_0$, which will be the order type of $\leq_0$ contrary to assumption.

$\phi[x, \alpha]$ is defined as "if f is a (set) function with domain an initial segment of $\leq_0$ containing x and having the property $f(y) = \{f(z) \mid z \leq_0 y\}$ for each y in its domain, then $f(x) = \alpha$". It is straightforward

to prove that exactly one such function f exists for each initial segment of $\leq_0$ (its extendability at limits in $\leq_0$ uses Replacement).

We have already seen that provision of this formula leads to a contradiction to our original assumption.

Corollary: The von Neumann cardinal $|A|$ exists for every set A.

Proof: There is a well-ordering of A, whose order type is an ordinal with the same cardinality of A. Either this is the smallest ordinal (in the inclusion order) with this property, in which case it is $|A|$ itself, or it has elements which have this property, among which there must be a smallest, which is $|A|$.

Theorem: V_α exists for each α.

Proof: Consider the smallest ordinal λ for which V_λ does not exist (it is obviously a limit ordinal if it exists).

Find a formula $\phi[\alpha, A]$ which says "$A = V_\alpha$" and we are done, because we can then define the set $\{F_\phi(\alpha) \mid \alpha \in \lambda\}$, and the union of this set will be V_λ contrary to assumption.

The formula $\phi[\alpha, A]$ says "there is a function f whose domain is an ordinal β such that $\alpha \in \beta$, and $f(0) = \emptyset$, $f(\gamma + 1) = \mathcal{P}(f(\gamma))$ if $\gamma + 1 \in \beta$, and $f(\mu) = \bigcup\{f(\gamma) \mid \gamma \in \mu\}$ for each limit ordinal $\mu \in \beta$, and $f(\alpha) = x$". The fact that there is a unique such function f for each $\beta < \lambda$ is readily shown: Replacement is used to show extendability of f at limit ordinals.

Zermelo set theory augmented with the Axioms of Replacement and Foundation is known as *ZFC* (Zermelo-Fraenkel set theory with Choice). This is the system of set theory which is most commonly used.

The Axiom of Foundation has been mentioned: we restate it.

Axiom of Foundation: For each set x, there is $y \in x$ such that $x \cap y = \emptyset$

The intention of this axiom is to assert that the membership relation restricted to any set is well-founded.

An informal way to explain the motivation of the axiom is that the sets are constructed in well-ordered stages. If x is any set, there must be a first

stage at which an element y of x is constructed (possibly more than one, but we select one). The idea then is that y, if it is constructed at a stage of positive index, must have all of its elements constructed at earlier stages, so none of them belong to x, so $x \cap y = \emptyset$. If y were constructed at stage 0 we would need to say more: but in our construction of ranks, stage 0 has no elements! Now we can observe that the Axiom of Restriction given earlier also implies Foundation for the same reason: stage 0 in the construction underlying the axiom of Restriction is $\mathbb{N}$, and the further remark to be made is that if the element y is constructed at stage 0 it is a natural number, and we can further choose y to be the smallest natural number belonging to x: its elements (whether we use the von Neumann or the Zermelo implementation of $\mathbb{N}$) are smaller natural numbers and so do not belong to x.

It is useful to note that the Axiom of Separation is almost redundant in the presence of Replacement.

Theorem: Zermelo set theory without Separation (and with the assertion that the empty set exists: this is part of the Axiom of Elementary Sets in the original formulation, but in more usual formulations it is deduced from Separation and the existence of any set at all, such as the one provided by Infinity), with the addition of the Axiom of Replacement, proves Separation.

Proof: Let $\phi[x]$ be a formula and let A be a set. If $\neg(\exists x : \phi[x])$ then $\{x \in A \mid \phi[x]\} = \emptyset$ exists. Otherwise, choose a such that $\phi[a]$ and define $\psi[x, y]$ as $\phi[x] \wedge y = x \vee \neg\phi[x] \wedge y = a$. Clearly ψ is a functional formula, so $\{y \mid (\exists x : x \in A \wedge \psi[x, y])\}$ exists, and this set is $\{x \in A \wedge \phi[x]\}$.

Although the Axiom of Replacement is sufficient to make the von Neumann definitions of cardinality and order type succeed, it is certainly not necessary. A weaker axiom with the same effect is the Axiom of Levels or the Axiom of Hierarchy, which can also be stated as the

Axiom of Beth Numbers: For every Scott ordinal α, $\beth_\alpha$ exists.

We define things in terms of Scott ordinals because we do not wish to presume that the von Neumann ordinal α exists; that is what we are trying to prove. A set of size $\beth_\alpha$ must be included in a rank V_β with $\beta \geq \alpha$, and the von Neumann ordinal α will be present in $V_{\beta+1}$. Notice that the Axiom of Rank plays an essential role in this argument: existence of large $\beth$ numbers

in the original Zermelo set theory does not have any effect on existence of von Neumann ordinals.

Another axiom which works is the stronger

Axiom of Beth Fixed Points: For every cardinal κ, there is a cardinal $\lambda > \kappa$ such that $\beth_\lambda = \lambda$.

A consequence of Foundation and Replacement which is often useful is the Mostowski Collapsing Lemma which we now present.

Review of definitions: If R is a relation, we define $R\text{“}A$ as $\{y \mid (\exists x \in A : x\,R\,y\}$. Thus $R^{-1}\text{“}A$ is $\{x \mid (\exists y \in A : x\,R\,y\}$. A relation R is *well-founded* iff for every subset A of $\mathtt{fld}(R)$ there is $a \in A$ such that $R^{-1}\text{“}\{a\}$ is empty (a is R-minimal). Notice that a nonempty well-ordering is not well-founded, but a strict well-ordering is.

Mostowski Collapsing Lemma: For every well-founded relation R, there is a unique function f with domain $\mathtt{fld}(R)$ such that

$$(\forall a \in \mathtt{fld}(R) : f(a) = \{f(b) \mid b\,R\,a\}).$$

Proof of Lemma: Define $\mathtt{cl}_R(a)$ for each $a \in \mathtt{fld}(R)$ as the intersection of all sets I such that $a \in I$ and $R^{-1}\text{“}I \subseteq I$: this can be thought of as the downward closure of $\{a\}$ under R.

Consider the set D of all elements a of $\mathtt{fld}(R)$ such that $R \cap \mathtt{cl}_R(a)^2$ satisfies the condition stated in the Lemma, that is, there is a unique function f_a with domain $\mathtt{cl}_R(a)$ such that $(\forall b \in \mathtt{cl}_R(a) : f_a(b) = \{f(c) \mid c\,R\,b\})$.

If $D = \mathtt{fld}(R)$ then observe that the set F of all such functions f_a exists by Replacement (the function f_a is a unique object associated with each $a \in \mathtt{fld}(R)$ in a way which can be defined by a formula) and $\bigcup F$ is the desired function f. To see this observe that if any two functions $f_a, f_b \in F$ both have c in their domain, the restrictions of each of f_a and f_b to $\mathtt{cl}_R(c)$ (which is a subset of both $\mathtt{cl}_R(a)$ and $\mathtt{cl}_R(b)$) must actually be the function f_c, so f_a and f_b agree at c, and $\bigcup F$ is a function (and certainly satisfies the stated conditions).

If $D \neq \mathtt{fld}(R)$ then there must be an R-minimal $d \in \mathtt{fld}(R) - D$. Each $c\,R\,d$ has an associated function f_c, If we extend the union of the

functions f_c for $c\,R\,d$ (which must be a function by the same argument above) to a function g with the additional value $g(d) = \{f_c(c) \mid c\,R\,d\}$, the function g satisfies the conditions to be f_d (and clearly is the only function which can do this), and so $d \in D$, which is a contradiction.

Theorem: For every set A, there is a well-founded relation R with unique function f associated to it by the Lemma such that f is the identity function on the field of R and A is in the range of f (so $f(A) = A$).

Proof: Suppose otherwise. Let B be a counterexample. We seek a further special counterexample C. If B has no elements which are counterexamples, then let $C = B$; otherwise let C be the $\in$-minimal counterexample belonging to B. In either case, C is a counterexample and none of its elements are counterexamples. With each $D \in C$, associate R_D, the intersection of all well-founded relations with D in their range satisfying the condition that the associated function f_D provided by the Mostowski Collapsing Lemma is the identity function. Take the union of the sets R_D and add the pair $\langle C, C\rangle$ to it as an element. This relation R_C satisfies the desired conditions, which is a contradiction.

This is a weird way of putting a proof of a useful result. For each set z, a relation R which is well-founded, has z in its field, and has the associated function f equal to the identity function on the field of R must actually be the restriction of the membership relation to $\mathtt{fld}(R)^2$. The field of the intersection of all such sets must be a transitive set containing z, and in fact the smallest one (because membership on any transitive set containing z satisfies the indicated conditions!). This set is called $\mathtt{TC}(\{x\})$, the transitive closure of $\{x\}$. ($\mathtt{TC}(x)$ would differ in not containing x as a member). There is an exercise which addresses proving the existence of this set in a different way.

An alternative pair of axioms to adjoin to Zermelo set theory which would tidy up such questions as existence of versions of the natural numbers and von Neumann order types of every well-ordering would be the combination of Foundation and the Mostowski Collapsing Lemma. This would correct technical problems with the original formulation of Zermelo set theory, but in a different way. Proving the existence of V_ω would be straightforward. One could not prove the existence of $V_{\omega\cdot 2}$, but one could prove the existence of the von Neumann order type of any set well-ordering: one could prove the existence of lots of sets for which one could not prove the existence of the

rank of the cumulative hierarchy to which one would expect them to belong. A world satisfying the Zermelo axioms and Foundation and Mostowski could be built in countably many stages: let H_0 be V_ω. Define H_{n+1} as the set of elements of ranges of Mostowski collapse functions on well-founded relations included in $H(n)^2$.

Exercises

1. This proof will use Replacement.

   In the usual axiom set it is rather more involved than it seems it ought to be to show that every set is a subset of a transitive set (this is easily shown in cumulative type theory or in Zermelo set theory with the rank axiom, but the usual formulation of Zermelo set theory or *ZFC* has the Foundation axiom, which is weaker).

   I give an outline of a proof which you need to complete (there are models in the notes for the proof).

   Let X be a set. We want to prove that there is a transitive set which contains X. The idea is to prove that the collection of sets $\{\bigcup^{\mathbf{n}}(X) \mid n \in \mathbb{N}\}$ exists. Then you can show that the union of this set is transitive and contains X as a subset.

   Fill in the details. To prove the existence of $\{\bigcup^{\mathbf{n}}(X) \mid n \in \mathbb{N}\}$ by Replacement you need a formula $\phi[n, x]$ which says "$x = \bigcup^{\mathbf{n}} X$". As I said, there are models for this in the notes.

   Why does it follow immediately from "X is a subset of a transitive set" that X is an element of some transitive set as well?

   Define the *transitive closure* $\mathtt{TC}(x)$ for any x as the intersection of all transitive sets including x as a subset: this set contains exactly the elements of x, elements of elements of x, elements of elements of elements of x, and so forth. It exists by this exercise. Note that $\mathtt{TC}(\{x\})$ contains x as an element in addition.

2. Prove using the axioms of Zermelo set theory and the axioms of Replacement and Foundation that every set x belongs to some rank V_α. You may use the result established in the section that V_α exists for any ordinal α. Hint: Suppose that there is a set x which belongs to no rank of the cumulative hierarchy. Consider an $\in$-minimal element of $\mathtt{TC}(\{x\})$ which does not belong to any rank V_α; of course you have to say why the hypotheses imply that there is such an object, and why bad things follow from this.

   You will note that this is a proof that our Axiom of Rank is a consequence of the axioms of *ZFC*. Of course, you cannot assume the Axiom of Rank in your argument.

3. This question is intended to address the question of just how weird a model of Zermelo set theory without the Axiom of Rank can be.

   Work in *ZFC*. Define a set A as *bounded* iff its transitive closure (defined in the first exercise) contains finitely many von Neumann natural numbers. We refer to the first von Neumann natural not in the transitive closure of A as the bound of A. Verify the following points:

   (a) If a and b are bounded, $\{a, b\}$ is bounded.

   (b) If A is bounded, $\mathcal{P}(A)$ is bounded (but the bound might go up by one – do you see why?), and $\bigcup A$ is bounded (with the same bound). You should also show that the bounds of the sets $\mathcal{P}^k(A)$ eventually increase with k.

   (c) The set of Zermelo natural numbers is bounded.

   (d) If the bound of A is n, the set $\mathcal{P}^{>n}(A)$ of all subsets of A with more than n elements is also bounded with bound n (note that this can be iterated).

   (e) Apply the points above to argue that the collection of all bounded sets in the universe of *ZFC* is a model of Zermelo set theory in which the set of von Neumann natural numbers does not exist, in which $\{\mathcal{P}^n(X) \mid n \in \mathbb{N}\}$ does not exist for any X, and in which there are sets at least as large as any set which exists in the universe of *ZFC* (for this last point you may assume without proof that for an infinite set A, $\mathcal{P}(A)$ is the same size as $\mathcal{P}^{>n}(A)$; of course I would enjoy it if you could prove this.) Hint: show that there is a bounded set at least as large as V_α for each α.

3.6.4 Transfinite Induction and Recursion

The reader may already have the idea that we have been engaging in transfinite arguments and constructions analogous to arguments by induction and constructions by recursion analogous to familiar arguments by induction and constructions by recursion on the natural numbers. In this section, we confirm this impression explicitly.

Transfinite Induction Theorem: Let $\phi[x]$ be any formula. If we can show that for any ordinal α, $(\forall \beta <_\Omega \alpha : \phi[\beta]) \rightarrow \phi[\alpha]$, it follows that for any ordinal α, $\phi[\alpha]$.

Proof: Suppose that for any ordinal α, $(\forall \beta <_\Omega \alpha : \phi[\beta]) \rightarrow \phi[\alpha]$, and further that for some ordinal γ, $\neg\phi[\gamma]$. We can assume that γ is the smallest such ordinal, since $\leq_\Omega$ is a "well-ordering", apart from failing to be a set. But then we have $(\forall \beta \leq_\Omega \gamma : \phi[\beta])$, from which we can deduce $\phi[\gamma]$ by hypothesis, which is a contradiction.

Notice that this theorem formally resembles strong induction on the natural numbers.

We give a form which looks a little bit more like the usual formulation of induction on the natural numbers.

Definition: A *limit ordinal* is an ordinal which is not zero and which is not a successor.

Transfinite Induction Theorem (three-case form): Suppose that $\phi[\alpha]$ is a formula which implies that α is an ordinal, and we can prove that

1. $\phi[0]$
2. $(\forall \alpha : \phi[\alpha] \rightarrow \phi[\alpha + 1])$
3. For each limit ordinal λ, $(\forall \beta <_\Omega \lambda : \phi[\beta]) \rightarrow \phi[\lambda]$.

It follows that $\phi[\alpha])$ holds for all ordinal α.

Proof: This follows easily from the previous form. The thing is to show that the hypotheses imply that for any ordinal α, $(\forall \beta \leq_\Omega \alpha : \phi[\beta]) \rightarrow \phi[\alpha]$. This is asserted for limit ordinals as one of the hypotheses. It is always true for $\alpha = 0$, vacuously. For successors $\alpha = \alpha' + 1$, $(\forall \beta \leq_\Omega \alpha' + 1 : \phi[\beta])$ implies $\phi[\alpha']$ which in turn implies $\phi[\alpha' + 1]$ by the hypotheses.

Now we introduce transfinite recursion. The most general form is analogous to course-of-values recursion on the natural numbers, in which the value of $f(n)$ is defined in terms of the entire restriction of f to $\{m \in \mathbb{N} \mid m < n\}$.

Transfinite Recursion Theorem: Suppose that we can uniformly define an operation G which acts on functions whose domain is an ordinal (G here has to be a class function notation, as it has to act on any function whose domain is an ordinal, and there is no set of all functions whose domains are ordinals). Then on any ordinal α there is a function F such that for each $\beta \leq_\Omega \alpha$ we have $F(\beta) = G(F\lceil\beta)$, and all such functions F satisfying this condition (but with possibly different ordinal domains) agree on the intersections of their domains.

Transfinite Recursion Theorem (three case form): Suppose we have defined a constant x and operations F and G (defined as class function notations). Then for any limit ordinal α there is a function H with domain α such that

1. $H(0) = x$
2. $H(\beta + 1) = F(H(\beta))$ for all $\beta < \alpha$
3. $H(\lambda) = G(\{H(\beta) \mid \beta < \lambda\})$ for each limit ordinal $\lambda < \alpha$. Note that $\{H(\beta) \mid \beta < \lambda\}$ exists by Replacement if $H(\beta)$ has successfully been defined for each $\beta < \lambda$.

Further, such functions H with domains distinct ordinals agree on the intersections of their domains. This theorem could be thought of as a kind of transfinite analogue of the Iteration Theorem.

The reader should recognize the definition of the ranks V_α of the cumulative hierarchy as an example of transfinite recursion (presented in the three-case form). Since the functions we construct in either form of transfinite recursion actually agree on common parts of their domains if they have different ordinals as their domains, we have in effect defined an operation on *all* ordinals in each case, though this operation cannot be realized by a single set function.

Inductions and recursions restricted to ordinals below a specific ordinal γ are readily handled using the general forms, by making the formula $\phi[\alpha]$ true for all $\alpha \geq_\Omega \gamma$ in the case of induction, and by defining the operations G

(or F and G) as taking default values (such as $\emptyset$) at functions whose domain is an ordinal $\geq_\Omega \alpha$ in the case of the first form of the transfinite recursion theorem, or at ordinals $\geq_\Omega \alpha$ or sets too big to be subsets of the range of $H\lceil\alpha$ in the case of the three-case form.

As an example of transfinite recursion, we present definitions of addition and multiplication of ordinal numbers.

Definition: We define $\alpha + \beta$ for ordinals α, β:

1. $\alpha + 0 = \alpha$
2. $\alpha + (\beta + 1) = (\alpha + \beta) + 1$
3. $\alpha + \lambda = \sup(\{\alpha + \beta \mid \beta < \lambda\})$, for λ limit.

Definition: We define $\alpha \cdot \beta$ for ordinals α, β:

1. $\alpha \cdot 0 = 0$
2. $\alpha \cdot (\beta + 1) = (\alpha \cdot \beta) + \alpha$
3. $\alpha \cdot \lambda = \sup(\{\alpha \cdot \beta \mid \beta < \lambda\})$, for λ limit.

Exercises

1. Demonstrate that addition of ordinals is not commutative by demonstrating that $1 + \omega \neq \omega + 1$.

   Demonstrate that multiplication of ordinals is not commutative by demonstrating that $2 \cdot \omega \neq \omega \cdot 2$.

2. Write a recursive definition of exponentiation of ordinals, modelled on the definition of multiplication above. Use your definition to compute 2^ω and ω^2. Describe a well-ordering of order type ω^2. See if you can describe a well-ordering of type ω^ω (it is fairly easy to construct one using familiar concepts, but I don't know that anyone will come up with it).

3. (Project!) Prove the Transfinite Recursion Theorem.

4. (Project!) Prove the three-case form of the Transfinite Recursion theorem, then show that the three-case form implies the original form.

3.7 The theory of infinite ordinal and cardinal numbers in untyped set theory

This section brings together basic results we will need about arithmetic of possibly infinite (transfinite) ordinal and cardinal numbers. Some of these results depend on the Axiom of Choice, which has a powerful simplifying effect on cardinal arithmetic. We'll try to indicate where this happens.

3.7.1 Transfinite ordinal arithmetic

It is useful to recall that for any ordinal α, $\alpha + 1 = \alpha \cup \{\alpha\}$: the successor operation for ordinal numbers is the same as that for the (von Neumann) natural numbers, which are of course precisely the finite ordinal numbers.

We give set theoretic definitions for ordinal addition and multiplication (which can be shown to be equivalent to the recursive definitions given above).

Definition: If $\leq_1$ is a well-ordering of order type α and $\leq_2$ is a well-ordering of order type β, then the set $\leq_1 \oplus \leq_2$ is defined as

$$\{\langle\langle c, i\rangle, \langle d, j\rangle\rangle \mid (c \leq_1 d \wedge i = 0 \wedge j = 0) \vee (c \leq_2 d \wedge i = 1 \wedge j = 1)$$
$$\vee (c \in \mathtt{fld}(\leq_1) \wedge d \in \mathtt{fld}(\leq_2) \wedge i = 0 \wedge j = 1)\}.$$

It is straightforward to determine that $\leq_1 \oplus \leq_2$ is a well-ordering and that its order type, which we call $\alpha + \beta$, is completely determined by α and β.

Notice that this is not the same use of $\oplus$ as the one above in the discussion of addition of natural numbers.

Note that what we are doing in effect is creating disjoint copies of $\leq_1$ and $\leq_2$ and putting the copy of $\leq_1$ before the copy of $\leq_2$.

Definition: If $\leq_1$ is a well-ordering of order type α and $\leq_2$ is a well-ordering of order type β, then the set $\leq_1 \otimes \leq_2$ is defined as

$$\{\langle\langle c, d\rangle, \langle e, f\rangle\rangle \mid c \leq_2 e \wedge d \in \mathtt{fld}(\leq_1) \wedge f \in \mathtt{fld}(\leq_1) \wedge (c = e \rightarrow d \leq_1 f)\}.$$

It is straightforward to determine that $\leq_1 \otimes \leq_2$ is a well-ordering and that its order type, which we call $\alpha \cdot \beta$, is completely determined by α and β.

Notice that this is not the same use of $\otimes$ as the one above in the discussion of multiplication of natural numbers.

Note that what we are doing in effect is creating disjoint copies of $\leq_1$ indexed by elements of the field of $\leq_2$ and putting the copies of $\leq_1$ in the order dictated by their indices in the field of $\leq_2$.

Note at once that these operations are not commutative. $1 + \omega = \omega$ is immediate from this definition, while $\omega + 1$ is the successor of ω. $2 \cdot \omega = \omega$, while $\omega \cdot 2 = \omega + \omega$. In exercises above, you are expected to use the recursive definitions of the same operations to verify these facts. Verifying that the two definitions are equivalent is likely to appear as an exercise below.

Addition does have identity 0 and multiplication has identity 1 (left and right). Multiplication has the zero property (left and right). Addition and multiplication of ordinals are both associative. $\alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma$, but the left associative property of multiplication over addition does not hold (this is easy: $(1 + 1) \cdot \omega \neq 1 \cdot \omega + 1 \cdot \omega$).

We give a

recursive definition of exponentiation of ordinal numbers: Of course a similar definition of exponentiation on natural numbers could be given (and is actually in effect included here). There is a set theoretical definition of exponentiation of ordinals as well, but it is a bit technical.

1. $\alpha^0 = 1$
2. $\alpha^{\beta+1} = \alpha^\beta \cdot \alpha$
3. $\alpha^\lambda = \sup(\{\alpha^\beta \mid \beta < \lambda\})$ for λ limit.

3.7.2 Transfinite cardinal arithmetic

In this section we will discuss basic properties of order, addition, multiplication and exponentiation of transfinite cardinals assuming the Axiom of Choice. We may have some things to say about what can be proved without Choice.

The Cantor-Schröder-Bernstein theorem proved above establishes that the order $\leq$ already defined on cardinal numbers is a partial order.

We briefly discuss the official definition of cardinals in a slightly different way.

Definition: An *initial ordinal* is an ordinal α such that for every $\beta < \alpha$, $|\beta| < |\alpha|$ (equivalently, for no $\beta < \alpha$ do we have $\beta \sim \alpha$). For any set A define $|A|$ as the unique initial ordinal such that $A \sim \alpha$. That there is such an ordinal we show as follows: by the Well-Ordering Theorem, there is a well-ordering $\leq_A$ with field A; for each such well-ordering there is an ordinal $\alpha = \mathtt{ot}(\leq_A)$ such that $(\subseteq \lceil \alpha) \approx <_A$; isomorphic relations have fields of the same size (because an isomorphism between two relations is a bijection between their fields, among other conditions), so $\alpha \sim A$; the smallest ordinal α isomorphic to a well-ordering with field A will be an initial ordinal equinumerous with A (if there were any ordinal β less than this α with $\alpha \sim \beta$, it would have $|\beta| = |A|$ whence one could define a well-ordering of A of type β contrary to choice of α).

Note that we will use different notation for one and the same object when it is considered as an ordinal and when it is considered as a cardinal. For example, ω, the first countable ordinal, is the same object as $\aleph_0$, the cardinality of all countably infinite sets (and for that matter both are the same as $\mathbb{N}$, the set of natural numbers), and similarly, ω_1, the first uncountable ordinal, and $\aleph_1$, the first uncountable cardinal, are the same object. Note though that if we used different implementations of cardinals and/or ordinals, these notations would have different referents, though theorems about cardinals and ordinals as such would tend to remain the same.

Theorem: The natural order on cardinals is a linear order.

Proof: What we need to show is that for any two sets A and B we can construct either an injection from A to B or an injection from B to

A. The well-ordering theorem allows us to select a well-ordering $\leq_A$ of A and a well-ordering $\leq_B$ of B. Now we know that there is either an isomorphism between $<_A$ and $<_B$, or an isomorphism between $<_A$ and some $(<_B)_b$ or an isomorphism between $<_B$ and some $(<_A)_a$. Now recall that an isomorphism is a bijection between the fields of the relations involved. Thus we have either a bijection from A to B, or a bijection from A to a subset of B, or a bijection from B to a subset of A, whence we either have an injection from A into B or an injection from B into A as required, so either $|A| \leq |B|$ or $|B| \leq |A|$.

Notice that the axiom of choice was used here. We will prove in the not too distant future that the assertion that the natural order on cardinals is a linear order implies the axiom of choice: these two assertions are exactly equivalent. This is perhaps somewhat surprising, since linear ordering of size seems a very natural assumption about sizes of sets.

Theorem: The natural order on cardinals is a well-ordering.

Proof: Let C be a set of cardinals. Our aim is to show that C has a smallest element in the natural order. Let $\leq$ be a well-ordering of the power set of the union of C (which is an ordinal larger as a set than any of the cardinals in C). Consider the set of all well-orderings of elements of C. Every well-ordering in this set will be isomorphic to some segment restriction of $\leq$. Consider the set of all $x \in \mathtt{fld}(\leq)$ such that $\leq_x$ is isomorphic to some well-ordering of some element of C: there must be a $\leq$-smallest element m of this set, and the order type of $(\leq)_m$ will be the smallest element of C (it has to be the order type of a well-ordering of some element of C, but moreover the order type of a shortest such well-ordering, which will be the element of C itself, since the element of C is a cardinal).

Theorem: There is a surjection from A onto B iff $|B| \leq |A|$ (and B is nonempty if A is).

Proof: If there is an injection f from B to A, then we can define a surjection from A to B as follows: choose $b \in B$; map each element of A to $f^{-1}(a)$ if this exists and to b otherwise. This will be a surjection. If B is empty we cannot choose b, but in this case A is empty and there is obviously a surjection.

If there is a surjection f from A onto B, there is a partition of A consisting of all the sets f"$\{a\}$ for $a \in A$. Let C be a choice set for this partition. Map each element b of B to the unique element of $C \subseteq A$ which is sent to b by f. This map is obviously an injection.

Definition: In set theory *without* Choice, we define

$$|A| \leq^* |B|$$

as holding iff there is a surjection from B onto A. In the light of the previous Theorem, there is no need for this notation if we assume Choice.

Definition (repeated from above): We define $\aleph_0$ as $|\mathbb{N}|$. Elements of $\aleph_0$ are called *countably infinite sets*, or simply *countable sets*.

Theorem: $\aleph_0 + 1 = \aleph_0 + \aleph_0 = \aleph_0 \cdot \aleph_0 = \aleph_0$. It is straightforward to define a bijection between $\mathbb{N}$ and $\mathbb{N} \times \mathbb{N}$. The bijections between the naturals and the even and odd numbers witness the second statement. The successor map witnesses the first statement.

Theorem: Every infinite set has a countable subset.

Proof: Let A be an infinite set. The inclusion order on the collection of all bijections from initial segments of $\mathbb{N}$ to A satisfies the conditions of Zorn's Lemma and so has a maximal element. If the maximal element had domain a proper initial segment of $\mathbb{N}$, then the set would be finite. So the maximal element is a bijection from $\mathbb{N}$ to a subset of A.

Theorem: For every infinite cardinal κ, $\kappa + 1 = \kappa$.

Proof: Let A be an infinite set. The inclusion order on the set of all bijections from B to $B \cup \{x\}$, where $B \cup \{x\} \subseteq A$ and $x \notin B$, satisfies the conditions of Zorn's Lemma and so has a maximal element. It is nonempty because A has a countable subset. If the maximal element is a map from B to $B \cup \{x\}$ and there is $y \in A - (B \cup \{x\})$, then affixing $\langle y, y \rangle$ to the map shows that the supposed maximal element was not maximal.

An easier proof of this uses the previous theorem. $\kappa = \lambda + \aleph_0$ for some λ, since a set of size κ has a countable subset. It follows that $\kappa = \lambda + \aleph_0 = \lambda + (\aleph_0 + 1) = (\lambda + \aleph_0) + 1 = \kappa + 1$.

Corollary: If n is finite and κ is an infinite cardinal then $\kappa + n = \kappa$.

Theorem: For every infinite cardinal κ, $\kappa + \kappa = \kappa$.

Proof: Let A be an infinite set. The set of pairs of bijections f and g with $\mathtt{dom}(f) = \mathtt{dom}(g) = \mathtt{rng}(f) \cup \mathtt{rng}(g) \subseteq A$ and $\mathtt{rng}(f) \cap \mathtt{rng}(g) = \emptyset$ can be partially ordered by componentwise inclusion:

$$(f, g) \leq (f', g') \leftrightarrow f \subseteq f' \wedge g \subseteq g'.$$

This partial order satisfies the hypotheses of Zorn's Lemma (verifying this is left as an exercise). It is nonempty because A has a countable subset. Suppose that a maximal such pair of bijections f, g in the componentwise inclusion order has been constructed. Let B be the common domain of f and g. If there is no countably infinite subset in $A - B$, then $A - B$ is finite and $|B| = |A|$ by a previous result and the result is proved: otherwise take a countable subset of $A - B$ and extend the supposedly maximal pair of maps to a larger one.

Alternative Proof: We prove by transfinite induction that for any ordinal $\alpha = \lambda + n$, where λ is 0 or limit and n is finite (every ordinal can be written in this way in one and only one way – exercise), we have $2 \cdot \alpha = \lambda + 2n$. Note that the arithmetic operations here are operations on *ordinals*. We prove this using three-case induction (in which λ always stands for a limit ordinal and n for a finite ordinal):

zero: $0 = 0 + 0$. 0=0+0; $2 \cdot 0 = 0 = 0 + 2 \cdot 0$.

successor: Let $\alpha = \lambda + n$. Suppose that $2 \cdot \alpha = 2 \cdot (\lambda + n) = \lambda + 2 \cdot n$. Then $2 \cdot (\alpha + 1) = 2 \cdot ((\lambda + n) + 1)[= 2 \cdot ((\lambda + (n + 1))] = 2 \cdot (\lambda + n) + 2 = (\lambda + 2 \cdot n) + 2 = \lambda + 2 \cdot (n + 1)$, verifying the claim for $\alpha + 1 = \lambda + (n + 1)$.

limit: Suppose that μ is limit and for every $\beta = \lambda + n < \mu$ we have $2 \cdot (\lambda + n) = \lambda + 2 \cdot n$. Then $2 \cdot \mu = \mathtt{sup}_{\lambda+n<\mu}(\lambda + 2 \cdot n)$. This supremum is less than or equal to μ because for any $\lambda + n < \mu$ we have $\lambda + 2 \cdot n < \mu$ as well. The supremum is greater than or equal to μ because any $\alpha < \mu$ is of the form $\lambda + n < \mu$, so $\alpha + 1$ is of the form $\lambda + (n + 1) < \mu$, and is strictly less than $2 \cdot (\alpha + 1) = \lambda + 2n + 2$, which is less than or equal to the supremum in question. So $2 \cdot \mu = 2 \cdot (\mu + 0) = \mu = \mu + 2 \cdot 0$ as desired.

So we have for any cardinal κ that $2 \cdot \kappa = \kappa$ by the Lemma, the finite part being zero. It is very important to note that this is ordinal multiplication. Now $2 \cdot \kappa$ is the order type of an order $\leq_1 \otimes \leq_2$, where the cardinality of the field of $\leq_1$ is 2 and the cardinality of the field of $\leq_2$ is κ. The field of this relation is $\mathtt{fld}(\leq_1) \times \mathtt{fld}(\leq_2)$ which is of the form $\{x\} \times \mathtt{fld}(\leq_2) \cup \{y\} \times \mathtt{fld}(\leq_2)$ where x and y are the two elements of the field of $\leq_1$. This set is of cardinality $\kappa + \kappa$ (here we mean cardinal addition). The field of a well-ordering of type κ is of course of size κ. Isomorphic well-orderings have fields of the same sizes, so the ordinal fact $2 \cdot \kappa = \kappa$ implies the cardinal fact $\kappa + \kappa = \kappa$.

Corollary: If $\lambda \leq \kappa$ and κ is an infinite cardinal then $\kappa + \lambda = \kappa$: note that $\kappa \leq \kappa + \lambda \leq \kappa + \kappa = \kappa$.

Theorem: For every infinite cardinal κ, $\kappa \cdot \kappa = \kappa$.

Proof: Let A be an infinite set. The inclusion order on bijections from $B \times B$ to B, where $B \subseteq A$, satisfies the conditions of Zorn's Lemma. It is nonempty because A has a countable subset. Now consider a maximal function in this order, mapping $B \times B$ to B. If $A - B$ contains no subset as large as B, then $|B| = |A|$ by the previous result and the result is proved. Otherwise, choose $B' \subseteq A - B$ with $|B'| = |B|$. It is then easy to see from assumptions about B and B' and the previous result that the map from $B \times B$ to B can be extended to a bijection from $(B \cup B') \times (B \cup B')$ to $B \cup B'$, contradicting the supposed maximality of the bijection.

Corollary: If $\lambda \leq \kappa$ and κ is an infinite cardinal then $\kappa \cdot \lambda = \kappa$.

The arithmetic of addition and multiplication of infinite cardinals is remarkably simple. This simplicity depends strongly on the presence of Choice.

We now introduce exponentiation of cardinals.

Definition: B^A is defined as the set of functions from the set A to the set B. $|A|^{|B|}$ is defined as $|A^B|$. It is a defect of this traditional notation that if κ and λ are considered as sets, the meaning of κ^λ (the set of functions from λ to κ) is the same size as but not the same object as the referent of κ^λ when κ and λ are considered as cardinals (the cardinality of the set just mentioned).

Observation: $2^{|A|} = |\mathcal{P}(A)|$, because there is a one to one correspondence between subsets $B \subseteq A$ and characteristic functions $\chi_B : A \to \{0,1\}$ such that $\chi_B(a) = 1$ iff $a \in B$ (every element of 2^A is a χ_B). Thus we have $2^{|A|} > |A|$ by the theorem of Cantor already proved.

We present some rules of exponentiation which should look familiar (though the generalizations require justification). These are combinatorial results which do not depend on Choice: writing out explicitly how to get bijections witnessing these equations of cardinality might be good exercises.

Rules of exponentiation:

1. $\kappa^1 = \kappa; \kappa^0 = 1; 1^\kappa = 1; 0^\kappa = 1$ if $\kappa = 0$, 0 otherwise.
2. $\kappa^{\lambda+\mu} = \kappa^\lambda \cdot \kappa^\mu$
3. $(\kappa^\lambda)^\mu = \kappa^{\lambda\cdot\mu}$
4. $(\kappa \cdot \lambda)^\mu = \kappa^\mu \cdot \lambda^\mu$

We prove a sample result combining our various tools.

Theorem (AC – do you see where?): For any infinite cardinal κ, $\kappa^\kappa = 2^\kappa$.

Proof: Let $|A| = \kappa$. κ^κ is the cardinality of A^A, which is a subset of $\mathcal{P}(A^2)$, which is the same size as 2^{A^2}, whose cardinality is $2^{\kappa^2} = 2^{\kappa\cdot\kappa} = 2^\kappa$, the size of the collection of subsets of A or functions from A to 2. On the other hand there are clearly no more functions from A to 2 than there are from A to A (an injection can easily be presented). So

$$\kappa^\kappa \leq 2^{\kappa^2} = 2^{\kappa\cdot\kappa} = 2^\kappa \leq \kappa^\kappa.$$

We now present some information about the structure of the entire system of cardinals. We know that for every cardinal κ, the cardinal $2^\kappa > \kappa$. It follows that there is a smallest cardinal greater than κ.

Definition: For any infinite cardinal κ, we define $\kappa+$ as the smallest cardinal greater than κ.

The cardinal $\kappa+$ coincides with the notion defined in the next definition, if the axiom of choice holds. Note that if we admit the possibility that the axiom of choice might fail, we need to use an alternative definition of cardinal number, such as the Scott definition, under which $|A|$ is the set of the form $\{B \sim A \mid B \in V_\alpha\}$ for the smallest α for which this set is nonempty.

Definition: For any infinite cardinal κ and set A of size κ, we define $\aleph(\kappa)$ as the supremum of the set of order types of well-orderings of subsets of A. This is known as the Hartogs aleph function. It is clear that it does not depend on the choice of A.

Lemma: If α is the order type of a well-ordering of a subset of an infinite set A, so is $\alpha + 1$.

Proof: A has a countable subset B; $B \setminus \{x\}$ is countable, so $A \setminus \{x\} = (A \setminus B) \cup (B \setminus \{x\}) \sim (A \setminus B) \cup B = A$. So $A \setminus \{x\}$ also has a well-ordering of order type α, to which x can be appended to give a well-ordering of order type $\alpha + 1$.

Theorem (not using AC): It is not the case that $|\aleph(\kappa)| \leq \kappa$. It follows immediately that $\aleph(\kappa)$ is an initial ordinal. But note that for the moment this does not mean that we identify $|\aleph(\kappa)|$ with $\aleph(\kappa)$, as we may be using a different cardinal implementation.

Proof of theorem: Let A be of cardinality κ. If $|\aleph(\kappa)| \leq \kappa$, then $\aleph(\kappa)$ is the order type of the field of a well-ordering of a subset of A. But then $\aleph(\kappa) + 1$ is also the order type of a well-ordering of a subset of A, and this contradicts the definition of $\aleph(\kappa)$: $\aleph(\kappa)$ must be the supremum of all order types of well-orderings of subsets of A.

Theorem (using AC): For any infinite cardinal κ, $\kappa+ = \aleph(\kappa)$.

Proof of theorem: Let A be of cardinality κ. By AC, A has a well-ordering, so $\aleph(\kappa) \geq \kappa$. By the previous theorem and the linearity of the natural order on cardinals, $\aleph(\kappa) > \kappa$. Now any ordinal strictly less than $\aleph(\kappa)$ is the order type of a well-ordering of a subset of A, so cannot be greater than κ, whence $\aleph(\kappa) = \kappa+$.

Theorem: The Axiom of Choice is true iff the natural order on cardinal numbers is a linear order.

Proof: We have already shown that if the axiom of choice is true, the natural order on cardinal numbers is a linear order.

Suppose that the natural number on cardinal numbers is a linear order. Let A be an infinite set of cardinality κ. We must have either $\kappa \leq \aleph(\kappa)$ or $\aleph(\kappa) \leq \kappa$. The latter is impossible by an earlier theorem. So we have

$\kappa \leq \aleph(\kappa)$, and any subset of a well-orderable set is well-orderable: there is a well-ordering on A because it is the same size as a subset of the set of order types of well-orderings of subsets of A, which is evidently well-orderable.

We now present some information about the structure of all cardinals (assuming Choice):

Definition: The notation $\aleph_0$ is already defined. We define $\aleph_{\alpha+1}$ as $\aleph_\alpha+$. We define α_λ, for λ limit, as $\mathtt{sup}(\{\aleph_\beta \mid \beta < \lambda\})$.

Observation: Under the axiom of choice, all cardinals are of the form $\aleph_\alpha$.

Definition: The notation $\beth_0$ is defined as $\aleph_0$. We define $\beth_{\alpha+1}$ as $2^{\beth_\alpha}$. We define $\beth_\lambda$, for λ limit, as $\mathtt{sup}(\{\beth_\beta \mid \beta < \lambda\})$.

Observation: Independently of Choice, the cardinals $\beth_\alpha$ are exactly the cardinalities of the ranks of the cumulative hierarchy with infinite index. $\beth_\alpha = |V_{\omega+\alpha}|$. For $\beta \geq \omega^2$, $\beth_\beta = |V_\beta|$. The Generalized Continuum Hypothesis is equivalent to the assertion that $\beth_\alpha = \aleph_\alpha$ for each α.

We introduce a further idea bearing on structure of cardinals.

Definition: The *cofinality* of a partial order $\leq$ is the infimum of the order types of unbounded well-ordered chains in $\leq$. Because the natural order on ordinals is a well-ordering, there will be a well-ordered chain in $\leq$ which is unbounded and has the cofinality of $\leq$ as its order type.

Definition: The cofinality of a cardinal κ, written $\mathtt{cf}(\kappa)$, is the cofinality of the inclusion order on κ itself.

Observation: The cofinality of $\mathtt{cf}(\kappa)$ is $\mathtt{cf}(\kappa)$; this is easily seen, because an unbounded well-ordered chain of minimal length in the inclusion order on $\mathtt{cf}(\kappa)$ will clearly be isomorphic to an unbounded well-ordered chain in the inclusion order on κ itself: if there were such a well-ordered chain of order type $< \mathtt{cf}(\kappa)$ in the inclusion order on $\mathtt{cf}(\kappa)$, there would be such a chain in the inclusion order on κ itself, contradicting the definition of $\mathtt{cf}(\kappa)$.

Observation: $\mathtt{cf}(\kappa)$ is an initial ordinal, and thus a cardinal. Suppose that $|\mathtt{cf}(\kappa)| < \mathtt{cf}(\kappa)$. Choose a bijection f from $|\mathtt{cf}(\kappa)|$ to $\mathtt{cf}(\kappa)$. For each $\alpha < |\mathtt{cf}(\kappa)|$, define a_α as $\mathtt{sup}_{\beta<\alpha}(\mathtt{max}(a_\beta + 1, f(\beta) + 1))$. The objects a_α are strictly increasing as α increases, and make up an unbounded [because any element $f(\alpha)$ of $\mathtt{cf}(\kappa)$ is less than or equal to a_α] well-ordered chain in the inclusion order on $\mathtt{cf}(\kappa)$ of order type $|\mathtt{cf}(\kappa)|$, whence we cannot have $|\mathtt{cf}(\kappa)| < \mathtt{cf}(\kappa)$. It might seem possible that $\mathtt{sup}_{\beta<\alpha}(\mathtt{max}(a_\beta + 1, f(\beta) + 1))$ might be $\mathtt{cf}(\kappa)$ for some $\alpha < |\mathtt{cf}(\kappa)|$: but this is ruled out because we would then have an unbounded well-ordered chain of order type $\alpha < \mathtt{cf}(\kappa)$ in $\mathtt{cf}(\kappa)$.

Observation on implementation-dependence: This discussion strongly depends on the use of the von Neumann ordinals and the use of initial ordinals to implement cardinals. An implementation-independent presentation which would work for different implementations of ordinals and cardinals is possible, and its character might be divined by taking a look at definitions of this notion in chapter 2.

Definition: A cardinal κ is *regular* iff $\mathtt{cf}(\kappa) = \kappa$. Note that cofinalities of cardinals are regular cardinals. A cardinal which is not regular is said to be *singular*.

Observation: $\aleph_0$ is regular.

Theorem (AC!): For any infinite cardinal κ, $\kappa+$ is regular.

Proof: Suppose otherwise. Then there would be an unbounded chain of order type $\lambda \leq \kappa$ in the inclusion order on $\kappa+$. Let C be this chain: C_α is the element of this chain, if any, such that the order type of the segment in C determined by C_α is α. We can assume that $C_0 \neq 0$. For each C_α choose a surjection f_α from κ onto C_α: this is possible (though only with the use of the Axiom of Choice) because any initial segment of an order of type $\kappa+$ is of cardinality $\leq \kappa$. We then define a surjection from $\kappa \times \lambda$ onto $\kappa+$: map $\langle \alpha, \beta \rangle$ to $f_\beta(\alpha)$. This means that $\kappa+ \leq |\kappa \times \lambda| = \kappa$, which is a contradiction.

Note that this theorem applies to the case $\kappa = \aleph_1$, or indeed any $\aleph_n$ for n finite.

Definition: A cardinal λ is said to be *strong limit* iff for each cardinal $\kappa < \lambda$ we have $2^\kappa < \lambda$. $\beth_\omega$ is the smallest uncountable strong limit cardinal. Note that $\beth_\omega$ is singular, having cofinality ω.

Definition: A regular strong limit cardinal is said to be *inaccessible*. We cannot give an example of one of these, as our current axioms cannot prove that there is one.

As our final major point in this section, we prove König's Theorem and explore some of its consequences.

Definition: We define infinite sums and products of cardinals. Let F be a function from an index set I whose range is a set of cardinals. We define $\Sigma_{i\in I}F(i)$ as $|\{\langle x, i\rangle \mid i \in I \wedge x \in F(i)\}|$, and we define $\Pi_{i\in I}F(i)$ as $|\{f \mid \mathtt{dom}(f) = I \wedge (\forall x \in I : f(x) \in F(i))\}|$.

It is important to note that the Axiom of Choice is required to establish the more general assertion that if A is a function from the index set I to sets such that $|A(i)| = F(i)$, that $\Sigma_{i\in I}F(i)$ is $|\{\langle x, i\rangle \mid i \in I \wedge x \in A(i)\}|$ and $\Pi_{i\in I}F(i)$ as $|\{f \mid \mathtt{dom}(f) = I \wedge (\forall x \in I : f(x) \in A(i))\}|$. The difficulty is that one needs not only bijections witnessing each assertion $A(i) \sim F(i)$, but also a uniform way to choose one such bijection for each i, in order to establish that the cardinalities of the last two expressions do not depend on the choice of the map A.

Note also that this definition would have to be rephrased if we were using a different definition of cardinality, under which the cardinality of the set was not itself a set of the same cardinality. Under a different definition of cardinality, it would be likely to be necessary to use the more general form which depends on the Axiom of Choice.

König's Theorem (depends on AC): Let F and G be functions with the same nonempty domain I whose ranges are sets of cardinals. Suppose further that $0 < F(i) < G(i)$ for each $i \in I$. It follows that $\Sigma_{i\in I}F(i) < \Pi_{i\in I}G(i)$.

Proof of König's Theorem: Suppose on the contrary that $\Sigma_{i\in I}F(i) \geq \Pi_{i\in I}G(i)$. It follows that there is a surjection H from $|\{\langle x, i\rangle \mid i \in I \wedge x \in F(i)\}|$ onto $|\{f \mid \mathtt{dom}(f) = I \wedge (\forall x \in I : f(x) \in G(i))\}|$. Each set A_i defined as $\{\langle x, i\rangle \mid x \in F(i)\}$ is of cardinality $F(i)$ and the

set π_i"H"$A_i \subseteq G(i)$ [where $\pi_i(f)$ is defined as $f(i)$] cannot cover $G(i)$ because $F(i) < G(i)$. Choose an element $k(i) \in G(i) - \pi_i$"H"A_i for each $i \in I$. The function k belongs to

$$|\{f \mid \mathtt{dom}(f) = I \wedge (\forall x \in I : f(x) \in G(i))\}|,$$

and cannot belong to the range of H (since by construction no element of any A_i can be mapped to k by H, and the union of the A_i's is the entire domain of H).

Consequences of König's Theorem: Cantor's theorem is a consequence: $\kappa = \Sigma_{i\in\kappa}1 < \Pi_{i\in I}2 = 2^\kappa$. Of course, Cantor's theorem can be proved without choice, so this is not an optimal proof of this result.

A very interesting corollary is that $\kappa^{\mathtt{cf}(\kappa)} > \kappa$ for any κ. Let $\lambda = \mathtt{cf}(\kappa)$ and let $\alpha_i < \kappa$ for $i \in \lambda$ be an unbounded strictly increasing sequence in κ. Then $\kappa = \Sigma_{i\in\lambda}\alpha_i < \Pi_{i\in\lambda}\kappa = \kappa^\lambda = \kappa^{\mathtt{cf}(\kappa)}$.

Now we can show that the cofinality of $2^{\aleph_0}$, the cardinality of the reals, is uncountable. If the cofinality of $2^{\aleph_0}$ is countable, then the preceding result establishes that $(2^{\aleph_0})^{\aleph_0} > 2^{\aleph_0}$. But $(2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0\cdot\aleph_0} = 2^{\aleph_0}$, so this is impossible. It turns out that this is basically the only provable limitation on which cardinal $2^{\aleph_0}$ can be.

3.7.3 Exercises

1. Let $\leq_{\mathbb{N}}$ be the usual order on the natural numbers. Use the operations $\oplus$ and $\otimes$ to describe orders of type $\omega + \omega$ and $\omega \cdot \omega = \omega^2$; draw illustrations of these orders sufficient to convince me that you know what they look like.

2. Prove that the two definitions of ordinal addition that you have been given (the one by transfinite recursion and the one using $\oplus$) actually agree. This should be a proof by transfinite induction.

3. Prove that a chain of injective functions in the inclusion order has an injective function as the union of its range (recall that for us a chain is actually a linear order; its range is the set which carries the linear order).

4. Verify that the componentwise inclusion order on pairs of bijections which appears in the first proof of $\kappa + \kappa = \kappa$ satisfies the hypotheses of Zorn's Lemma.

5. Verify rules 2 and 4 of exponentiation. Rule 3 is especially tricky, and you will receive additional credit (and praise) if you can prove it. Remember that sets B^A between which you are building bijections are sets of functions, and your bijections need to involve clever definitions of functions taking functions to other functions.

6. Explain why the Well-Ordering Theorem implies the axiom of choice (in the presence of the other axioms).

7. Show that for any cardinal κ, if $\lambda \leq \kappa$ and $\{\alpha_\beta\}_{\beta<\lambda}$ is an increasing unbounded sequence of length λ in κ, then $\Sigma_{i\in\lambda}\alpha_i = \kappa$. Hint: you can map the standard set whose cardinality is $\Sigma_{i\in\lambda}\alpha_i = \kappa$ into $\kappa \times \lambda$ (this is straightforward), which is a set of size κ (why?); the trick then is to see how to map κ injectively into this set, which will complete the proof. Hint: consider the sets $\alpha_{\beta+1} - \alpha_\beta$.

8. Prove that $(\beth_\omega)^{\aleph_0} = 2^{\beth_\omega}$. Hint: there is a standard set of size $\beth_\omega$ that you can think of, namely, the union of the iterated power sets of the natural numbers. Consider how you can use a sequence of elements of this set to approximate an arbitrary subset of this set.

9. (Project, entirely optional) A set theoretical definition of ordinal exponentiation β^α can be given. One expects this to have something to do with functions from a set with an order of type α on it to a set with an order of type β on it. In fact, this is true with a subtle modification. If $\mathtt{ot}(\leq_1) = \alpha$ and $\mathtt{ot}(\leq_2) = \beta$, the order $\leq_2^{\leq_1}$ has domain the set of functions from $\mathtt{fld}(\leq_1)$ to $\mathtt{fld}(\leq_2)$ which are equal to 0 at all but finitely many elements of $\mathtt{fld}(\leq_1)$. An order statement $f \leq_2^{\leq_1} g$ is evaluated by considering the largest value x in $\mathtt{fld}(\leq_1)$ at which the functions f and g disagree (there is a largest such value because f and g have the value 0 except at a finite number of inputs), and returning the truth value of $f(x) \leq_2 g(x)$.

   The project is to verify that this definition is equivalent to the recursive one given above. I imagine it is rather difficult.

3.8 Logically regimented set constructions and the definition of L

The reader should be aware of the correlation between propositional logic (the logic of "and", "or", and "not") and the Boolean algebra of sets (the logic of intersection, union and complement (relative to a fixed universe in the context of untyped set theory). In this section, we extend this idea to include the operations of the logic of quantifiers. We will use this in our development of Gödel's constructible universe.

Definition: A "predicate set" over a domain D is a collection of elements of D^∞ (defined as the set of natural-number-indexed sequences of elements of D which are eventually constant) which has a *predicate order* in a sense we now define: a subset X of D^∞ has predicate order $n \in \mathbb{N}$ iff for each $f \in X$, for every $g \in D^\infty$, if $f\lceil n = g\lceil n$, then $g \in X$.

The intention is that a predicate set is in all cases a set of sequences of elements of the domain indexed by all natural numbers, but if the predicate takes n arguments, then whether a sequence belongs to the associated predicate set depends only on the first n terms of the sequence (those with indices $< n$, since indexing starts at 0). Notice that a predicate set with predicate order n is also of predicate order m for all $m \geq n$.

Basic constructions of predicate sets: If P and Q are predicate sets of predicate order n, then $P^c = D^\infty \setminus P$ and $P \cap Q$ are predicate sets of predicate order n. This gives us support for the propositional operations of conjunction and negation, and so for all the operations of propositional logic.

For any $f \in D^{\mathbb{N}}$, define $f_{i,j}$ as $(f\lceil(\mathbb{N} - \{i, j\}) \cup \{\langle i, f(j)\rangle, \langle j, f(i)\rangle\}$. For any predicate set P of order n, $P_{i,j} = \{f_{i,j} \mid f \in P\}$ is a predicate set of predicate order the maximum of n, i, j.

For any predicate set P, define $\exists P$ as $\{\langle 0, d\rangle \cup f\lceil\mathbb{N}^+ \mid f \in P \wedge d \in D\}$. Define $\exists_i P$ as $(\exists P_{0,i})_{0,i}$. Define $\forall P$ as $(\exists P^c)^c$ and $\forall_i P$ as $(\exists_i P^c)^c$. These operations implement quantification on predicate expressions. If P is a predicate set of predicate order n, each of these sets is also a predicate set of predicate order n.

For any unary predicate P, define $[P]$ as $\{f \in D^\infty \mid P(f(0))\}$. This is clearly a predicate set of predicate order 1. For any logical relation R, define $[R]$ as $\{f \in D^\infty \mid f(0)\, R\, f(1)\}$. This is clearly a predicate set of predicate order 2.

Representations of propositions by predicate sets:

atomic formulas: Define $[Px_i]$ as $[P]_{0,i}$. Define $[x_i\, R\, x_j]$ as $([R]_{0,i})_{1,j}$, when $i \neq j$ and $i \neq 1$ and $j \neq 0$.. If $i \neq j$ and $i = 1$ and $j \neq 0$, define it as $([R]_{0,i})_{0,j}$. If $i \neq j$ and $i \neq 1$ and $j = 0$, define it as $([R]_{0,i})_{ij}$. If $i = 1$ and $j = 0$ define it as $[R]_{01}$. If $i = j$, define it as $\exists_{i+1}([x_i\, R\, x_{i+1} \wedge x_i = x_{i+1}])$.]

propositional operations: Define $[\neg\phi]$ as $[\phi]^c$. Define $[\phi \wedge \psi]$ as $[\phi] \cap [\psi]$. For any other propositional connectives, generate such definitions by redefining the connectives in terms of conjunction and negation.

quantifiers: Define $[(\forall x_i \in D : \phi)]$ as $\forall_i([\phi])$. Define $[(\exists x_i \in D : \phi)]$ as $\exists_i([\phi])$.

The point is that for any proposition ϕ, this procedure will create the set $[\phi]$ of all sequences f such that if each variable x_i is assigned the value $f(i)$, the proposition ϕ is assigned the truth value "true". This could be modified to allow different domains for different variables, by defining our universe as the collection of functions f such that for each $i \in \mathbb{N}$, $f(i) \in D(i)$, where D is a function from natural numbers intended to take each i to the set to which the variable x_i is to be bounded. This would require care in the use of the operators $\cdot_{ij}$, as they would transpose not only values but intended domains.

We can strengthen our position further by representing formulas ϕ as sets themselves.

predicate and relation symbols: For any set x, we allow $\langle 0, x\rangle$ to be an atomic unary predicate symbol and $\langle 1, x\rangle$ to be an atomic binary relation symbol. We provide that $\langle 0, 0\rangle$ represents the predicate `set` of sethood (if atoms are considered) and that for each $x \in D$, $\langle 0, \{x\}\rangle$ represents the predicate $P_x(y)$ defined as $y = x$, and $\langle 1, 0\rangle$ and $\langle 1, 1\rangle$ represent $=$ and $\in$, respectively. We could also provide constants, but we note that a constant c can always be handled by introducing a predicate C such that Cx_i means $x_i = c$, and we have arranged to be able

to do this by providing for each $x \in D$ a predicate true exactly of x. It does simplify things that the only "nouns" in the language we implement here are the variables x_i for natural numbers i. We could introduce other predicates, but for our immediate purposes we will not need to do this.

atomic formulas: We let $\langle 2, P^*, n\rangle$ represent Px_n, where P^* represents the logical predicate P. We let $\langle 3, R^*, m, n\rangle$ represent $x_m\, R\, x_n$, where R^* represents the logical relation R.

propositional logic: We let $\langle 4, \phi^*, \psi^*\rangle$ represent $\phi \wedge \psi$, and $\langle 5, \phi^*\rangle$ represent $\neg\phi$, where ϕ^* represents ϕ and ψ^* represents ψ.

quantifiers: We let $\langle 6, \phi^*, i\rangle$ represent $(\exists x_i.\phi)$, where ϕ is represented by ϕ^*.

other logical operations: One may add more clauses for further logical operations and quantifiers or view them as always abbreviating constructions using the operations given, which are adequate.

We can then associate with every set which is an expression according to the definition just given a predicate set which it is intended to represent, given intended representations for each symbol $\langle 0, x\rangle$ and $\langle 1, x\rangle$. The notation $\mathtt{ref}$ below everywhere abbreviates $\mathtt{ref}_{\mathcal{U},\mathcal{R}}$, where $\mathcal{U}, \mathcal{R}$ are specific functions explained in the first clause.

predicate and relation symbols: We define $\mathtt{ref}(\langle 0, x\rangle)$ as a predicate set $\mathcal{U}(x)$ of order 1 for each x we use as a unary predicate symbol in our language. We define $\mathtt{ref}(\langle 1, x\rangle)$ as a predicate set $\mathcal{R}(x)$ of order 2 for each x we use as a binary predicate symbol in our language. We stipulate that $\mathcal{U}(0)$ is $\{f \in D^\infty \mid \mathtt{set}(f(0))\}$ [if we make use of atoms, which for the most part we will not] and that for each $x \in D$, $\mathcal{U}(\{x\}) = \{f \in D^\infty \mid f(0) = x\}$ [so that we have a predicate which picks out each individual member of D, whether we can actually characterize it with a formula or not], and $\mathcal{R}(0)$ is $\{f \in D^\infty \mid f(0) = f(1)\}$ and $R(1)$ is $\{f \in D^\infty \mid f(0) \in f(1)\}$

atomic formulas: Define $\mathtt{ref}(\langle 2, P, n\rangle)$ as $\mathtt{ref}(P)_{0,n}$, if $\pi_1(P) = 0$ and $\mathtt{ref}(P)$ is defined.

For any natural numbers m, n, p, define $(mn)p$ as n if $p = m$, m if $p = n$, and otherwise as p.

Define $\mathtt{ref}(\langle 3, R, m, n\rangle$ as $((\mathtt{ref}(R))_{0,m})_{(0,m)1,n}$, when $m \neq n$ and further $\pi_1(R) = 1$ and $\mathtt{ref}(R)$ is defined. Just swapping m with 0 and n with 1 does not always work.

Define $\mathtt{ref}(\langle 3, R, m, m\rangle$ as

$$\mathtt{ref}(\langle 6, \langle 4, \langle 3, R, m, m+1\rangle \langle 3, \langle 1, 0\rangle, m, m+1\rangle\rangle, m+1\rangle)$$

when $\pi_1(R) = 1$ and $\mathtt{ref}(R)$ is defined. The idea here is that $x_m \, R \, x_m$ is equivalent to $(\exists x_{m+1} : x_m \, R \, x_{m+1} \wedge x_m = x_{m+1})$, and that is what that nasty expression does.

propositional logic: Define $\mathtt{ref}(\langle 4, \phi, \psi\rangle$ as $\mathtt{ref}(\phi) \cap \mathtt{ref}(\psi)$ where $\mathtt{ref}(\phi)$ and $\mathtt{ref}(\psi)$ are defined. Define $\mathtt{ref}(\langle 5, \phi\rangle)$ as $(\mathtt{ref}(\phi))^c$, where $\mathtt{ref}(\phi)$ is defined.

quantifiers: Define $\mathtt{ref}(\langle 6, \phi, i\rangle)$ as $\exists_i(\mathtt{ref}(\phi))$, when $\mathtt{ref}(\phi)$ is defined.

other logical operations: One may add more clauses for further logical operations and quantifiers or view them as always abbreviating constructions using the operations given, which are adequate.

The added power here is that we have imported all the formal sentences of our logical language into our mathematical universe, and have assigned meanings to all sentences, subject to the condition that all quantifiers are restricted to the particular domain set D (or to sets $D(i)$ in the alternative approach we sketched).

We can further establish that both the collection of sets representing logical formulas and the function $\mathtt{ref}_{U,R}$ just defined are actually sets.

Definition: Let D be a set. Let $\mathcal{U}$ and $\mathcal{R}$ be functions with range included in D^∞ and satsifying further conditions described above. A $\mathcal{U}, \mathcal{R}$-*formula-inductive set* is a set I with the following closure properties:

1. $\langle 0, P\rangle$ is in I iff $P \in \mathtt{dom}(\mathcal{U})$. $\langle 1, R\rangle$ is in I iff $R \in \mathtt{dom}(\mathcal{R})$.
2. $\langle 2, P, n\rangle$ is in I iff n is a natural number, $\pi_1(P) = 0$, and $\pi_2(P) \in \mathtt{dom}(\mathcal{U})$. $\langle 3, R, m, n\rangle$ is in I iff m, n are natural numbers, $\pi_1(R) = 1$ and $\pi_2(R) \in \mathtt{dom}(\mathcal{R})$.

3. $\langle 4, P, Q\rangle$ and $\langle 5, P\rangle$ are elements of I if P and Q belong to I.
4. $\langle 6, P, n\rangle$ belongs to I iff $P \in I$ and n is a natural number.

Definition: $\mathcal{L}_{\mathcal{U},\mathcal{R}}$ is defined as the intersection of all $\mathcal{U},\mathcal{R}$-formula inductive sets. The existence of a set including the natural numbers and the domains of $\mathcal{U}$ and $\mathcal{R}$ and closed under pairing will provide such an inductive set to start with: a limit rank of the cumulative hierarchy containing both of these sets does the trick, and in fact less is needed (just a rank of infinite index including those sets) if one uses a different ordered pair definition. Note that the countable set V_ω suffices if the domains of $\mathcal{U}$ and $\mathcal{R}$ are finite (and so may harmlessly be supposed inhabited by natural numbers). The letter $\mathcal{L}$ should suggest "language": these are the sentences of a formal language, internalized as objects of our set theory.

Definition: Where $D, \mathcal{U}, \mathcal{R}$ are as above, a $\mathtt{ref}_{\mathcal{U},\mathcal{R}}$-inductive set is a set I which is a relation with domain $L_{\mathcal{U},\mathcal{R}}$ and range D^∞ and has the following closure properties:

1. $\langle\langle 0, x\rangle, \mathcal{U}(x)\rangle$ belongs to I for each x in the domain of $\mathcal{U}$. $\langle\langle 1, x\rangle, \mathcal{R}(x)\rangle$ belongs to I for each x in the domain of $\mathcal{R}$.
2. $\langle\langle 2, P, n\rangle, X_{0,n}\rangle$ belongs to I if $\langle P, X\rangle$ belongs to I and $\pi_1(P) = 1$.
3. $\left\langle\langle 3, R, m, n\rangle, (X_{0,m})_{(0m)1,n}\right\rangle$ belongs to I if $\langle R, X\rangle$ belongs to I, $\pi_1(R) = 1$, and $m \neq n$.
4. $\langle\langle 3, R, m, m\rangle, X\rangle$ belongs to I if

$$\langle\langle 6, \langle 4, \langle 3, R, m, m+1\rangle, \langle 3, \langle 1, 0\rangle, m, m+1\rangle\rangle, m+1\rangle, X\rangle$$

   belongs to I.
5. $\langle\langle 4, P, Q\rangle, X \cap Y\rangle$ is in I if $\langle P, X\rangle$ and $\langle Q, Y\rangle$ are in I.
6. $\langle\langle 5, P\rangle, X^c\rangle$ is in I if $\langle P, X\rangle$ is in I.
7. $\langle\langle 6, P, n\rangle, \exists_i X\rangle$ is in I if $\langle P, X\rangle$ is in I.

Definition: The relation $\mathtt{ref}_{\mathcal{U},\mathcal{R}}$ is defined as the intersection of all $\mathtt{ref}_{\mathcal{U},\mathcal{R}}$-inductive sets. Notice that $\mathcal{L}_{\mathcal{U},\mathcal{R}} \times D^\infty$ is such an inductive set. It might be an instructive exercise to prove that this is a function with domain $\mathcal{L}_{\mathcal{U},\mathcal{R}}$. Once it is seen to be a function, it is seen to satisfy the conditions in its earlier informal definition.

Notice that we can characterize a formula in $L_{\mathcal{U},\mathcal{R}}$ as simply true if its image under $\mathtt{ref}_{\mathcal{U},\mathcal{R}}$ is the universal predicate set D^∞, and false if its image is $\emptyset$. it is very important to observe that we have only defined this notion of truth for formulas in which every formula is bounded in the set D. We could further adapt this to allow each variable x_i to be bounded in a set D_i as suggested above, but in any case all quantifiers must be bounded in sets.

We can now define the constructible universe L of Gödel. It is the "union" (not a set) of a sequence of ranks indexed by the cumulative hierarchy, but not the same ranks as in the case of the universe V.

For any set D, and the minimal functions $\mathcal{U}$ and $\mathcal{R}$ defined exactly as above with no additional predicates, define $\mathtt{Def}(D)$ as the collection of all sets E such that $\{f \in D^\infty \mid f(0) \in E\}$ is in the range of $\mathtt{ref}_{\mathcal{U},\mathcal{R}}$. In other words $\mathtt{Def}(D)$ is the collection of all subsets of D definable in first-order logic with all quantifiers bounded in D, with no primitive logical notions available other than membership, equality, and each element of D considered as a constant [and in addition sethood if atoms are admitted].

We then define

1. $L_0 = \emptyset$ [or the set of all atoms if atoms are present and make up a set].
2. $L_{\alpha+1} = \mathtt{Def}(L_\alpha)$
3. $L_\lambda = \bigcup\{L_\beta \mid \beta < \lambda\}$ for λ limit.

The whole exertion of this section was in showing that $\mathtt{Def}$ can indeed be defined internally to our set theory. We say that a set is *constructible* if it is an element of some L_α, and we refer to the collection of all constructible sets (which is certainly not a set) as the *constructible universe* L. We will assume that there are no atoms unless we specifically state otherwise.

It is perhaps worth observing that the existence of $L_{\alpha+1}$ follows from the existence of L_α and the axioms of Zermelo set theory, but the existence of L_λ in general requires Replacement (or at least the axiom of hierarchy).

3.8.1 Defining the well-ordering on L

We now examine what sets we can construct in L.

Theorem: If ϕ is a formula in which every quantifier is bounded in a set in L, then $\{x \in L_\alpha \mid \phi\}$ belongs to some L_β.

Proof of Theorem: Quantifiers bounded to specific sets are definable because individual objects (such as the specific sets to which the quantifiers are bounded) have predicates which pick them out.

It is not necessarily the case that $\beta = \alpha + 1$: what we can say is that if γ is the maximum of all the ordinals δ such that some object named in ϕ belongs to L_δ, then $\{x \in L_\alpha \mid \phi\}$ belongs to $L_{\max(\alpha,\delta)+1}$, because $\{x \in L_\alpha \mid \phi\} = \{x \in L_{\max(\alpha,\delta)+1} \mid x \in L_\alpha \wedge \phi\}$, and this clearly belongs to $\mathtt{Def}(L_{\max(\alpha,\delta)})$.

It is important to observe that every finite subset of L_α belongs to $L_{\alpha+1}$. Iterated application of this fact tells us that every bijection from a natural number to a set A exists in $L_{\alpha+3}$ if A is in L_α, and so the collection of finite subsets of L_α is in something like $L_{\alpha+4}$, definable as the collection of subsets of L_α which are the same size as an element of ω.

It is important to note that there is no *prima facie* reason to believe that all countably infinite subsets of a given set of L_α are found in L_β for any $\beta > \alpha$ whatsoever.

If $D \in L_\alpha$, every pair of an element of D and a natural number exists in $L_{\alpha+2}$, and every element of D^∞ exists in $L_{\alpha+3}$, and the set D^∞ itself exists in $\alpha + 4$ (the test on a countable sequence for whether it is in D^∞ has to do with whether its range is finite, and one can already identify finite subsets of D at the level of $L_{\alpha+4}$. Nothing hangs on getting these small finite numbers exactly right.

For any set x, we can define $\mathtt{TC}(\{x\})$ as the collection of all y such that there is a finite sequence s with domain $n+1$ such that $s_0 = y$, $s_i \in s_{i+1}$ for each $i < n$, and $s(n) = x$. If $x \in L_\alpha$, all terms of this sequence will be in L_α because L_α is a transitive set, and it will be definable in $L_{\alpha+4}$ or so.

We claim that $x = \mathtt{Def}(D)$ is definable by a formula with every quantifier bounded in L. This is rather involved. We can certainly describe D^∞. We can describe the minimal sets $\mathcal{U}$ and $\mathcal{R}$.

That an object x belongs to $\mathcal{L}_{\mathcal{U},\mathcal{R}}$ is equivalent to the existence of a finite subset of $\mathtt{TC}(\{x\})$ with certain closure properties: define the parse tree of

x as the intersection of all sets T which contain x and have the following closure properties:

1. if $y = \langle 2, P, n\rangle \in T$ then $P \in T$
2. if $y = \langle 3, R, m, n\rangle \in T$ then $R \in T$
3. if $y = \langle 4, P, Q\rangle \in T$ then $P, Q \in T$
4. if $y = \langle 5, P\rangle \in T$ then $P \in T$
5. if $y = \langle 6, P, n\rangle \in T$ then $P \in T$

The parse tree of any x must be finite (there cannot be an infinite descending sequence of "subterms" of x by Foundation). We say that $x \in L_{\mathcal{U},\mathcal{R}}$ iff every element of the parse tree of x which does not have first projection between 2 and 5 inclusive is either $\langle 0, 0\rangle$ (in case we are considering atoms) or a $\langle 0, \{x\}\rangle$ for $x \in D$, or $\langle 1, 0\rangle$ or $\langle 1, 1\rangle$.

That a pair $\langle x, P\rangle$, with $x \in \mathcal{L}_{\mathcal{U},\mathcal{R}}$ and $P \in D^\infty$, belongs to $\mathtt{ref}_{\mathcal{U},\mathcal{R}}$ is similarly witnessed by a finite set, the intersection of all sets T which contain $\langle x, P\rangle$ and have closure conditions

1. If $\langle\langle 4, P, Q\rangle, A\rangle \in T$, then there are B and C such that $\langle P, B\rangle \in T$ and $\langle Q, C\rangle \in T$ and $A = B \cap C$.
2. If $\langle\langle 5, P\rangle, A\rangle \in T$, then $\langle P, A^c\rangle \in T$.
3. If $\langle\langle 6, P, n\rangle, A\rangle \in T$, then there is B such that $\langle P, B\rangle \in T$ and $\exists_n(B) = A$.
4. If $\langle\langle 2, P, n\rangle, A\rangle \in T$ then $\langle P, A_{0,n}\rangle \in T$
5. If $\langle\langle 3, R, m, n\rangle, A\rangle \in T$ then there is B such that $\langle P, B\rangle \in T$ and $A = (B_{0,m})_{(0m)1,n}$.
6. If $\langle\langle 3, R, m, m\rangle, A\rangle \in T$ then the mutant translation of it given above also belongs to T.

If $\langle\langle 0, X\rangle, A\rangle$ or $\langle\langle 1, X\rangle, A\rangle$ belong to the intersection of all such T, then the value of A is required to be determined by the intended semantics in the obvious way.

The minimal such set will be finite, with first coordinates restricted to the parse tree of x (mod additions caused by the repeated arguments clause for relations, which will only be applied finitely many times).

Now, since we can define the formula $x = \texttt{Def}(D)$ in such a way that we can use it in the definitions of sets in L, this means that we can say what it means for a set to be an L_α: we can assert that there is a sequence indexed by an ordinal in which each term is the image under `Def` of the previous term (if there is a previous term) or the union of all previous terms if there is no previous term.

Now we can define a well-ordering on L (global on the entire universe). Certainly L_0 is well-ordered (uniquely). Suppose for each $\beta < \alpha$ that we have well-ordered L_β, and that further that the order that we have defined on each L_γ with $\gamma < \beta$ is a segment restriction of the order on L_β. It follows immediately that we have a well-ordering on L_α if α is limit, and if the order on each L_β for $\beta < \alpha$ is defined in a uniform way which can be described in language bounded to an $L_{\alpha+i}$ for a small finite i; it remains to show how to get a well-ordering on L_α if $\alpha = \beta + 1$. We have an order on L_β already and we know that we want each element of L_β to appear before all elements of $L_\alpha - L_\beta$ in this order. So, to decide which order two elements of $L_\alpha - L_\beta$ are to be placed in, we appeal to an order on $\mathcal{L}_{\mathcal{U},\mathcal{R}}$, and we place x before y iff x is defined in $\texttt{Def}(L_\beta)$ by a formula which appears before any formula defining y. Since we can define formulas and their references in L in a way which supports definition of sets in L, all that remains is to define an order on the formulas in $\mathcal{L}_{\mathcal{U},\mathcal{R}}$. Lexicographic order suffices: order first by the natural number initial in the formula, then recursively by the same order on simpler formulas for subformulas, by numerical order for numeral components, and (tricky last point) by the order already defined for L_β for constants in L_β appearing as components of $\mathcal{U}$. Now, the successor case shows us how the order on $L_{\beta+1}$ is determined in a uniform way from the order on L_β, which allows us to meet the condition stated above which is required at limit ordinals: we need to say not only that we have defined L_γ for each $\gamma < \alpha$ limit, but also that each $L_{\gamma+1}$ is determined by L_γ as discussed in the successor case (which is expressible in suitably bounded language), in addition to the assertion already made that for $\delta < \gamma < \alpha$, the order on L_δ is a segment restriction of the order on L_γ.

Now we can define a Hilbert symbol $(\epsilon x : \phi[x])$ for any formula ϕ: $y = (\epsilon x : \phi[x])$ means that $\phi[y]$ is true and y is the least object in the well-order on L for which this is true, or that there is no x such that $\phi[x]$ and $y = \emptyset$.

This works even if ϕ is an unbounded formula.

3.8.2 L satisfies the axioms of ZFC

We show that the universe L of constructible sets satisfies the axioms of ZFC. We remark that though we have no collection L, we can make sense of "x is in L" as meaning "There is an ordinal α such that $x \in L_\alpha$".

We consider the axioms one by one.

Extensionality: We want to show that if x and y are in L and they have the same elements which belong to L, then they are the same set. The crucial point here is that L is transitive: if x is in L, there is a first L_α to which x belongs, which must be an $L_{\beta+1} = \mathtt{Def}(L_\beta)$. Any $y \in x$ then clearly belongs to L_β and so is in L. Note that not only L but each L_α is transitive for just these reasons. Thus if x and y are in L, and have the same elements belonging to L, then they have the same elements in the real world V (the collection of all sets of our set theory) and so they are equal.

Pairing: If $x \in L_\alpha$ and $y \in L_\beta$ then $\{x, y\} \in L_{\mathtt{max}(\alpha,\beta)+1}$ for obvious reasons.

Union: If $x \in L_\alpha$, then the union of x is definable in $L_{\alpha+1}$ as usual, since all elements of x and elements of elements of x also belong to L_α, which is a transitive set as noted above.

Infinity: $\omega \in L$.

Foundation: This follows directly from Foundation in V and the fact that L is transitive (the fact that L is transitive isn't even needed, but it makes it easier).

Power Set: This is the first hard case. Suppose $A \in L_\alpha$. The collection $\mathcal{P}(A) \cap L$ is a set by Separation in the real world. Let $\phi(x, \beta)$ mean '"β is the smallest ordinal such that $x \in L_\beta$. This is a functional formula. By Replacement in the real world V, the collection of all ordinals β such that $\phi(x, \beta)$ holds for some $x \in \mathcal{P}(A) \cap L$ is a set. The union of this set is an ordinal γ, and $\mathcal{P}(A) \cap L$ appears in $L_{\gamma+1}$, defined as the collection of all elements of L_γ which are included in A as a subset. $\mathcal{P}(A) \cap L$, once it is seen to be an element of L, witnesses the truth of the Axiom of Power Set with all quantifiers restricted to L. Notice that many or most subsets of A which are in V may not ever appear in L at all: this power set may be quite impoverished when viewed as it were from the outside.

Separation: To prove separation (and replacement) we need to give some definitions and prove a lemma.

Definition: Where M is a subcollection of the universe, what we mean by saying that a formula ϕ is true in M should be formalized (we have already been talking about this informally). We read $M \models \phi$ as "M says that ϕ is true" or "ϕ is true in M".

1. $M \models x \in y$ or $M \models x = y$ means the same thing as $x \in y$ or $x = y$ just in case $x \in M$ and $y \in M$.
2. $M \models \neg\phi$ just in case $\neg(M \models \phi)$. $M \models (\phi \wedge \psi)$ just in clase $(M \models \phi) \wedge (M \models \psi)$.
3. $M \models (\exists x : \phi)$ just in case $(\exists x \in M : M \models \phi)$. Notice the important restriction of the quantifier here.

Notice that this definition works equally well if M is a set A or if M is a "collection" defined by a formula, such as L.

Definition: If $M \subseteq N$ are collections, we say that M agrees with N about $\phi[x_1, \ldots, x_n]$, where the x_i's are all the free variables in ϕ, iff $M \models \phi[t_1, \ldots, t_n]$ if and only if $N \models \phi[t_1, \ldots, t_n]$ for all $t_1, \ldots, t_n \in M$.

Lemma: For every formula ϕ and ordinal β, there is an L_α with $\alpha > \beta$ which agrees with L about ϕ.

Proof of Lemma: Fix a formula ϕ.

For any formula $\psi[x, x_1, \ldots, x_n]$, define $\beta_\phi(x_1, \ldots, x_n)$ as the first ordinal such that $L_{\beta_\phi(x_1,\ldots,x_n)}$ contains a t such that $L \models \phi[t, x_1, \ldots, x_n]$, or 0 otherwise. Define $\mathtt{cl}_\phi(A)$ for any set $A \in L$ as the supremum of all $\beta_\psi(x_1, \ldots, x_n)$ for ψ a subformula of ϕ and values of x_i's taken from A. Define $\mathtt{cl}^0_\phi(A)$ as A and $\mathtt{cl}^{n+1}_\phi(A)$ as $\mathtt{cl}_\phi(\mathtt{cl}^n_\phi(A))$. The various closure sets exist by applications of Replacement. It is important to notice that we are dealing with only finitely many formulas at a time: we can define $\beta_\phi(x_1, \ldots, x_n)$ for concretely given formulas (finitely many of them) but we cannot uniformly define what it means for an arbitrary formula to be true in L inside L (or even inside V!).

It is straightforward to establish that $\mathtt{cl}^\omega(A)$, the union of all $\mathtt{cl}^n_\phi(A)$'s agrees with L about ϕ when all values of free variables in ϕ are taken from $\mathtt{cl}^\omega(A)$: this is proved by induction on the

definition of $\models$, and relies on the fact that for any subformula of ϕ and any choice of free variables from $\mathtt{cl}^\omega(A)$ other than x, we have arranged for a witness to $(\exists x : \phi)$ to exist in $\mathtt{cl}^\omega(A)$ (in the next $\mathtt{cl}^n_\phi(A)$ after the first one which contains all the values assigned to free variables) just in case it exists in L.

Now the desired L_α is $\mathtt{cl}^\omega(L_\beta)$.

Now we can complete the proof of separation. Let A be a set in L_α and let ϕ be a formula (in which there may be unbounded quantifiers over L). Let L_γ be a level of L above L_α which agrees with L about ϕ (notice that this works if free variables appear in ϕ, too). The set $\{x \in A \mid \phi\}$ in the sense proper to L then appears in $L_{\gamma+1}$, defined as the collection of all elements of L_γ which belong to L and satisfy ϕ as localized to L_γ.

Replacement: Suppose $A \in L_\alpha$ and suppose $\phi[x, y]$ is a functional formula. Find an L_γ with $\gamma > \alpha$ which agrees with L about ϕ. The collection of y such that L says that for some $x \in A$, $\phi[x, y]$ is definable in $L_{\gamma+1}$ since L_γ agrees with L about ϕ (and so contains all the needed images!).

Choice: We showed in the previous section that L sees a well-ordering of every set. So if P is a partition in L, well-order $\bigcup P$ using that partition and let C be the collection of first elements in that order of elements of P.

3.8.3 L satisfies GCH

We first observe that the cardinality of L_α is the same as the cardinality of α, for each infinite ordinal α. We argue for this by transfinite induction. This is as true in L as it is true in V.

Clearly $|L_\omega| = \omega$: L_ω is the union of countably many finite sets L_n.

Suppose that L_β has the same cardinality as β. $L_{\beta+1} = \mathtt{Def}(L_\beta)$ obviously has cardinality at least that of β, since it includes β as a subset. We show that it has cardinality at most that of β: each element of $\mathtt{Def}(\beta)$ is associated with one or more finite length expressions in a language with $|\beta|$ symbols, so the size of $\mathtt{Def}(\beta)$ is bounded by the size of the collection of such expressions. The collection of such expressions of length n is at most $|\beta|^n = |\beta|$, and the collection of all such expressions of all lengths has size at most $|\beta| \cdot \aleph_0 = |\beta|$.

Suppose that L_β is of size $|\beta|$ for each $\beta < \lambda$ limit. The union is of size no more than $|\lambda| \cdot |\lambda| = \lambda$ (it is no larger than a disjoint union of copies of all the L_β's). It is clearly of size at least $|\lambda|$ since it contains all elements of λ.

Let $A \in L_\alpha$ and suppose that $|A| = |\alpha|$ according to L (we can arrange this for any cardinal $|\alpha|$ in the sense of L by considering $A = \alpha \in L_{\alpha+1}$). Each $B \in \mathcal{P}(A) \cap L$ appears first in some L_γ. We claim (and must prove below) that $|\alpha| = |\gamma|$ and moreover that this is true in L. We note first that this will prove our result: every subset of A must then appear in $L_{(|\alpha|+)_L}$, which L itself sees as having cardinality $|\alpha|+$ (the ordinal $(|\alpha|+)_L$ which L thinks is the next cardinal after $|\alpha|$ might be of the same cardinality as α as far as V is concerned, because L might be missing some bijections!). So L sees the cardinality of $\mathcal{P}(A)$ (which it sees as $2^{|\alpha|}$) as bounded by $|\alpha|+$, and on the other hand the cardinality of $\mathcal{P}(A)$ must be at least $|\alpha|^+$ by Cantor's theorem.

It remains to establish our claim.

Recall that we defined a Hilbert symbol $(\epsilon x : \phi[x])$ above, as the L-first object x such that $\phi[x]$, or else 0 if there is no such object. We define a closure in a way similar to the way we defined a closure in the Lemma proving Separation above, though this one will be a smaller set. We begin with all elements of L_α and all elements of the set B (which has no more than $|\alpha|$ elements). We go through countably many steps. At each step we add $(\epsilon x : \phi[x, t_1, \ldots, t_n])$, with ϕ being interpreted in the sense of L_γ, and all t_i's added at previous stages. Notice that here we can refer to all formulas uniformly: we know how to determine whether any formula at all is true with its quantifiers restricted to L_γ and its parameters taken from objects

present at the previous stage, using our logically regimented set construction machinery. After ω steps, we have built a structure M which agrees with L_γ about *every* formula: this is established by an induction on the definition of $\models$.

Further, this structure M is of size $|\alpha|$, for the same sort of reason that each L_α is of size $|\alpha|$: it is built using finite strings of symbols taken from a previous stage which may be supposed of size $|\alpha|$ in an inductive argument.

Now we take a Mostowski collapse of M. M just thinks it is an L_α (because it has delusions that it is L_γ). But membership on M is extensional (if two elements of M are distinct, we might find that some elements of these elements of M are not in M, but because L_γ knows that they are different, some element of their symmetric difference (describable by a Hilbert symbol) is in M) and well-founded (because it is a subrelation of true membership on L_γ, which is well-founded) so we can collapse M using a Mostowski collapse.

Each element of the collapsed set M^* actually has exactly the members the theory of M^* says it has, and in fact M^* is an L_β: we can see this by induction on $\delta < \beta$. Let δ be the first ordinal such that L_δ as M^* sees it is not the real L_δ. This δ cannot be 0. It cannot be a limit λ because it is then the union of all the things M^* sees as L_χ for $\chi < \delta$, and this really will be L_δ because M^* correctly identifies all the earlier ones. Finally, if M^* identifies L_χ correctly, it also identifies $\mathtt{Def}(L_\chi)$ correctly (because the elements of $\mathtt{Def}(L_\chi)$ are defined using finite expressions built up from symbols taken from L_χ) so the first bad L_δ cannot be a successor $L_{\chi+1}$ either! M^*, since it thinks it is L_γ, sees itself as either $\mathtt{Def}(L_\delta)$ for some δ or a union of L_δ's, and since it has a correct understanding of what sets are levels of L, this means that it is itself a level of L.

So we have an L_β of size $|\alpha|$ (it is the same size as M) which contains all elements of B. $B \in L_\beta$ because it is actually defined in exactly the same way it was defined in L_γ! But $|\beta| = |\alpha|$, so the first possible γ must also have been of this cardinality (it is a final weird consequence of this argument that if we chose the first possible γ that in fact $\gamma = \beta$, and nothing much really happened in the collapse).

3.8.4 Final remarks about L

If we accepted the universe of constructible sets as the universe of sets, we would thereby answer almost all questions about set theory. Why do we not believe that L is the universe?

I intend to add more (but still brief) discussion of this question.

Exercises

1. Read the section on L and send me any remarks you have about typos, points of confusion, and so forth.

2. Determine which L_α contains the set of all finite functions with domain and range included in ω (this is a bookkeeping problem).

3. If the set D belongs to L_α, determine the smallest finite n for which it can be shown that $D^\infty \in L_{\alpha+n}$ (this is again bookkeeping, like the previous problem). You might want to write out actual definitions of typical elements of D^∞ to see how this works.

   Explain why I cannot expect to be able to define $D^{\mathbb{N}}$ in L (if D is infinite).

4. Prove that for each finite n, $L_n = V_n$. (This is straightforward).

   Prove that $L_\omega = V_\omega$ (give the brief justification on the basis of what we have already shown).

   Now prove that $L_{\omega+1} \neq V_{\omega+1}$ (something we have shown recently will show this immediately).

   It might be that $L_\alpha = V_\alpha$ will have *only* the finite ordinals and ω as solutions for α. If $V \neq L$, there are very simple situations under which this will happen (describe such a situation).

   If $V = L$, there is a next α above ω such that $L_\alpha = V_\alpha$, and you should be able to explain what this value of α is and why on the basis of things shown recently.

5. (this might be rather evil) If D is a transitive set, demonstrate that the set $X = \{A \in D \mid |A| = 1\}$ belongs to $\mathtt{Def}(D)$ by showing how to construct a subset Y of D^∞ using the axioms of cylindrical algebra such that the set $\{y(0) \mid y \in Y\} = X$. Explain why it is necessary

for me to assume that D is a transitive set in order for you to be able to define this. My mental model of this is that you will need to carry out a series of definitions of subsets of D^∞ using the cylindrical algebra operations based on the defining formula of X.

3.9 Theories with proper classes

In this section, we outline approaches to foundations basically similar to what we have done so far, modified to allow us to speak of large collections like the Russell class as objects. The key idea is that the very large collections (which we call "classes") cannot themselves be members of classes.

We present the axioms of a theory of this kind. The primitive predicates of the theory are equality and membership.

The empty class; definitions of atom and class: There is a distinguished object $\emptyset$ with no elements, which we call the empty class. Objects with no elements are called atoms. Objects with elements and the empty class are called classes.

Axiom of Extensionality: Classes with the same elements are equal.

Definition: We say that x is a *set* iff $(\exists y : x \in y)$: elements are sets. A class which is not a set is called a *proper class*.

Axiom of Comprehension: For any formula $\phi[x]$, there is a class $\{x \in V \mid \phi[x]\}$ such that for each a, $a \in \{x \in V \mid \phi[x]\}$ if and only if a is a set and $\phi[a]$. We define V as $\{x \in V \mid x = x\}$. V is the class of all sets.

Axiom of Elementary Sets: $\emptyset$ is a set. For any sets x, y, $\{x, y\}$ is a set.

Axiom of Power Set: For any set x, $\mathcal{P}(x)$ is a set.

Axiom of Union: For any set x, $\bigcup x$ is a set.

Axiom of Infinity: ω is a set.

Axiom of Limitation of Size: For any class A, A is a proper class if and only if there is a class bijection from A to V.

Axiom of Foundation: Each class has an element disjoint from itself.

We summarize why the sets of this theory satisfy the axioms of ZFC. Extensionality for sets (in its strong form if we assume that there are no atoms) follows immediately from Extensionality for classes.

The empty class is a set by Elementary Sets. Pairing, Power Set, Union, and Infinity for sets are explicitly provided. Foundation is explicitly provided.

Choice holds, strangely enough, by Limitation of Size. The class of all von Neumann ordinals which are sets exists by Comprehension, is obviously a von Neumann ordinal, and cannot be a set on pain of the Burali-Forti paradox. Thus by Limitation of Size the class of all von Neumann ordinals can be placed in one-to-one correspondence with the universe. The class of von Neumann ordinals can be well-ordered, and from this well-ordering we obtain a well-ordering of the universe V, which gives us Choice (in a very strong form, in fact).

Replacement holds, because if there is a functional relation from a set A to a class C, the functional relation can be implemented as a class bijection, and since A is not of the same cardinality as V, and $C = f``A$ cannot have larger cardinality than A, it follows that C is not of the same cardinality as V, and so is a set. Separation we have seen follows from Replacement and the existence of the empty set.

The axiom of limitation of size gives a different and stronger approach to what properties have extensions which are sets than the approach implicit in separation (that a property has an extension which is a set if its extension is included in something already known to be a set). The idea is that the common property of sets is that they are smaller than the universe.

The theory that we have described, which is called Morse-Kelley set theory, is somewhat stronger than ZFC. To get a theory of essentially the same strength as ZFC, restrict the Axiom of Comprehension to apply only to formulas ϕ in which every quantifier is restricted to a class.

3.9.1 Pocket set theory, or, who said mathematicians don't have a sense of humor?

Pocket set theory is a theory with sets and classes which doesn't allow very large collections (or does it?). It is based on the observation that the only cardinals which "occur in nature" are $\aleph_0$ and c, the cardinality of the set of natural numbers and the cardinality of the reals. Its axiomatics are also just plain funny.

The primitive predicates of pocket set theory are equality and membership. General objects of the theory are called *classes*. For simplicity we rule out atoms.

Axiom of Extensionality: Classes with the same elements are equal.

Definition: We say that x is a *set* iff $(\exists y : x \in y)$: elements are sets. A class which is not a set is called a *proper class*.

Axiom of Comprehension: For any formula $\phi[x]$, there is a class $\{x \in V \mid \phi[x]\}$ such that for each a, $a \in \{x \in V \mid \phi[x]\}$ if and only if a is a set and $\phi[a]$. We define V as $\{x \in V \mid x = x\}$. V is the class of all sets.

Thus far the theory is almost the same as the theory with sets and classes given above. We do not postulate the axiom of pairing (we will be able to prove it) but we define unordered pairs, ordered pairs and class bijections as usual (though as yet we do not know that there are any).

Definition: A class A is *infinite* iff there is a class bijection from A to a proper subset of A. We say that two classes are the same size iff there is a class bijection from one of the classes to the other.

Axiom of Infinite Sets: There is an infinite set, and all infinite sets are the same size.

Axiom of Proper Classes: All proper classes are the same size, and no proper class is the same size as a set.

We now prove a series of theorems, getting at the end to the point where we can see the shape of the world of pocket set theory.

Theorem: The Russell class $R = \{x \in V \mid x \notin x\}$ is a proper class.

Proof: This is familiar.

Theorem: The empty class $\{x \in V \mid x \notin x\}$ is a set.

Proof: Otherwise the empty class is a proper class and so is the same size as the Russell class, so the Russell class is empty. Let I be an infinite set. Then $\{I\} = \{x \in V \mid x \in I\}$ is a set, because it is clearly not the same size as the Russell class. It is not infinite, because it clearly (having one element) cannot be the same size as a proper subclass of itself. But $\{I\}$ is then clearly a member of the Russell class (as it is not an element of itself, being distinct from its infinite sole element), which is a contradiction.

Theorem: For any set x, $\{x\} = \{y \in V \mid y = x\}$ is a set.

Proof: Suppose that $\{x\}$ is a proper class for some set x. Thus the Russell class is the same size as $\{x\}$ and has exactly one element. Let I be an infinite set. $\{I, \emptyset\} = \{y \in V \mid y = I \vee y = \emptyset\}$ is then a set, because there clearly cannot be a class bijection from this class to $\{x\}$. $\{I, \emptyset\}$ clearly belongs to the Russell class, as it cannot be equal to either of its elements by reason of size. $\emptyset$ is a set and also belongs to the Russell class. But the Russell class is supposed to have exactly one element, so we have arrived at a contradiction.

Theorem: For any sets x, y, $\{x, y\} = \{x \in V \mid z = x \vee z = y\}$ is a set. This is readily seen, as $\emptyset$, $\{\emptyset\}$ and $\{\{\emptyset\}\}$ are all elements of the Russell class, so it cannot be placed into a one to one correspondence with $\{x, y\}$.

Theorem: For any sets x, y the ordered pair $\langle x, y\rangle$ is a set. For any logical relation R, there is a class $\{\langle x, y\rangle \in V \mid x\, R\, y\}$ implementing R. Thus, if there is a logically describable bijection between two classes, there is actually a class bijection between them.

Theorem: The class of von Neumann ordinals is a proper class. It is obviously a proper class von Neumann ordinal, which we will call ω_1.

Proof: The reasons for this are familiar.

Theorem: The universe V can be well-ordered.

Proof: By the axiom of proper classes, V is the same size as ω_1.

Theorem: There is an infinite ordinal.

Proof: An infinite set I will be the same size as a subclass of ω_1, which will be the same size as an initial segment of ω_1, which will be an ordinal. This ordinal is a set because it is the same size as I and it is infinite because I is infinite.

infinite set ordinals discussed: We define ω as the first infinite ordinal, which we know to be a set. All ordinals α with $\omega \leq \alpha < \omega_1$ are sets (because they belong to ω_1 and the same size as ω because they are infinite. We see that ω is the familiar ordinal of that name, and we see that ω_1, being the first uncountable ordinal, is also the familiar ordinal of that name.

real numbers implemented: Natural numbers are represented as elements of ω as usual. Positive rationals can be represented as pairs of positive natural numbers (and so are sets). Reals can then be defined as initial segments of the positive rationals as elsewhere in these notes: these are countable classes and thus sets. The class of reals is the same size as the power class of ω for the usual reasons. The power class of ω is larger than ω for the usual reasons. Because it is larger than ω it is not a set and so is the same size as ω_1. Thus we obtain not only Choice from our version of Limitation of Size (the axiom of proper classes can be viewed thus) but also the Continuum Hypothesis.

There are two ways to view this. We may suppose that we have a system in which all collections are very small (they being no larger than the system of real numbers) or we may suppose that our view is that the class of reals is very large.

3.10 Forcing

We tend to see how forcing works through the lens of concepts developed (somewhat independenty? I'm not sure of the history) for the model theory of constructive logic. This sometimes causes our terminology to be nonstandard. We also take a different approach to implementation of membership and equality which avoids mutual recursion between the definitions of forcing of membership and equality conditions, though this does have some other costs.

We also need a result from model theory. By an inner model, we mean a model whose membership relation is a subset of the true membership relation.

Theorem: Let T be any set theory with a transitive inner model which is a set (by inner model, I mean that the membership relation of the model is actually the membership relation of the real world). Then T has a countable transitive model (ctm for short).

Proof: Start with any transitive inner model of T. Put a well-ordering on the model, and define $(\epsilon x.\phi[x])$ as the first object t such that $\phi[t]$ if there is one, and as $\emptyset$ otherwise (or some suitable default object in the model of T). Start with the empty set as the first stage; at each stage add all referents of Hilbert symbols $(\epsilon x.\phi(x, t_1, \ldots, t_n))$ where the values t_i of the free variables are taken from previous stages, and all quantifiers are understood as bounded in T. Go through ω stages. The resulting structure is countable (because every object in it is represented by a finite string written in a finite or countable alphabet) and its membership relation is well-founded and extensional but not necessarily transitive. Take a Mostowski collapse and you get a model which again satisfies the same sentences but which is a countable transitive inner model of the original theory.

We can prove the existence of such a model of any theory which has a transitive inner model which is a set. We cannot prove in ZFC itself that ZFC has a transitive inner model which is a set, and so we cannot prove that ZFC has a ctm. If there is an inaccessible cardinal, the existence of such a transitive inner model follows, and so the existence of a ctm of ZFC (because if κ is inaccessible, V_κ is a transitive inner model of ZFC). We also get such a model if we have a magic oracle which tells us what sentences are true in ZFC and a global order on

ZFC; these tools would allow us to define the Hilbert symbols and build a ctm of full ZFC. Also note that we can definitely build a ctm of Zermelo set theory, because $V^{\omega \cdot 2}$ is a transitive inner model.

We now work inside a ctm of ZFC (however obtained) or inside a ctm of a suitable weaker theory.

Now for the logical concepts.

I'm going to outline a semantics originally developed for constructive logic to motivate forcing. Suppose we have a set P with an order $\leq$ on it. The elements of P are called *conditions* and represent states of knowledge. $p \leq q$ means that we have more information in condition q than we do in condition p (but compatible with condition p). We say that conditions p and q are *compatible* iff there is a condition r such that $p \leq r$ and $q \leq r$; conditions which are incompatible represent states of knowledge which are inconsistent with one another. We resist the usual habit of writing the partial order in the other sense: most workers use $\geq$ where we use $\leq$.

splitting property: We require that P have the *splitting property*: for every $p \in P$, there are q and r such that $p \leq q$, $p \leq r$, and q and r are incompatible

Definition (truth value sets): We call a subset τ of P a *truth value set* if it satisfies two conditions:

1. For every $p \in \tau$, for every $q \geq p$, $q \in \tau$.
2. If $(\forall q \geq p : \exists r \geq q : r \in \tau)$, then $p \in \tau$.

forcing sets: We associate with each proposition ϕ a truth value set $[\phi]$: we refer to $[\phi]$ as the set of conditions which force ϕ (so $p \in [\phi]$ can be read "p forces ϕ" or less formally, "p asserts ϕ"). The idea is that $[\phi]$ is the set of stages of knowledge at which we are able to determine that ϕ is true. The first condition in the definition of truth value sets is thus motivated: if we believe ϕ at the stage p, then we will still believe ϕ at any stage $q \geq p$.

some clauses in the recursive definition of forcing sets: We regard our operations of propositional and first-order logic as being defined in terms of negation, conjunction, and the universal quantifier.

1. The forcing set $[\neg\phi]$ is defined as $\{p \in P \mid (\forall q \geq p : q \notin [\phi])\}$. A condition forces $\neg\phi$ if neither that condition nor any stronger condition forces ϕ.

   The definition of forcing sets of negations motivates the second condition in the definition of truth value sets: this condition is precisely what is needed to ensure that $[\neg\neg\phi] = [\phi]$.

2. The forcing set $[\phi \wedge \psi]$ is defined as $[\phi] \cap [\psi]$: a condition forces $\phi \wedge \psi$ iff it forces both ϕ and ψ.

3. The forcing set $[(\forall x \in N : \phi[x])]$ is defined as $\bigcap_{x \in N}[\phi[x]]$, where N is the set of names of objects in our domain of discourse. The reason that we speak of names of elements of the domain of discourse rather than elements themselves is that equations between names will have nontrivial forcing sets: there may be names for elements of our domain which under some conditions are names for the same object and under some conditions are names for different objects. A condition forces $(\forall x \in N : \phi[x])$ iff it forces $\phi[x]$ for every name.

The definitions of forcing sets for disjunctions and existential quantifiers follow from the definitions given and the natural definitions of disjunction and the existential quantifier in terms of negation, conjunction, and the universal quantifier, but they may be a bit unexpected. $[\phi \vee \psi]$ is not $[\phi] \cup [\psi]$, which might not be a truth value set at all, but the smallest truth value set which contains this set. For example, the forcing set $[\phi \vee \neg\phi]$ is the entire set P, though certainly there will be conditions at which we will not have decided whether ϕ or $\neg\phi$ holds: however, above any condition $q \geq p$ there is a condition r at which either ϕ or $\neg\phi$ is asserted.

The semantics for $[(\exists x : \phi[x])]$ then follow from the definitions already given, but again might seem to require comment: p asserts $(\exists x \in N : \phi[x])$ just in case for every condition $q \geq p$ there is $r \geq q$ such that r asserts $\phi[t]$ for some name t of an object in our domain of discourse, but it does not necessarily follow from $p \in [(\exists x \in N : \phi[x])]$ that there is any name t such that $p \in [\phi[t]]$.

We will now introduce our domain of names. We build a subclass of names of sets in our ctm of ZFC (or whatever theory). We build the names in a hierarchy reminiscent of others we have seen.

1. $N_0 = \emptyset$

2. $N_{\alpha+1} = \mathcal{P}(N_\alpha \times P)$

3. $N_\lambda = \bigcup_{\beta<\lambda} N_\beta$ for λ limit.

So a name is a relation between names (of lower rank) and conditions. The intention is that if $(x, p) \in y$ (x and y being names) that $p \in [x \in y]$. Things are trickier than that, though.

We define the rank of a name as the smallest α such that the name belongs to $N_{\alpha+1}$.

Note on restrictions on names which would tidy up this treatment

The following additional restrictions on names would clear up many difficulties below.

1. Define x_p (differently from below, without recursion, but for the names we construct with these restrictions the operation will be the same) as the set of all $(y, q) \in x$ such that $q \geq p$.

2. Require that if $(x, p) \in y \in N$ and $q \geq p$, then $x_p = x$ and $(x_q, q) \in y$ as well. Notice that $x_p = x$ ensures that if $(u, r) \in x$ we have $r \geq p$, and this further applies to u, to first projections of elements of u, and so forth.

3. Require that if $(x, p) \notin y$ then there must be $q \geq p$ such that for all $r \geq q$, $(x_r, r) \notin y$: this ensures that the set of p such that $(x_p, p) \in y$ is a truth value set.

   The actual effect of this note on the definition above is to restrict $N_{\alpha+1}$ to be not the entire set $\mathcal{P}(N_\alpha \times P)$ but the subset of this set determined by the restrictions just stated.

The standard approach involves a mutual recursion between definitions of $[x \in y]$ and $[x = y]$. It is sound, and even fairly easy to believe, but it is hard to demonstrate convincingly that one is not reasoning in a circle. We take a different approach.

reducing names under a condition: For any name x and condition p, we define x_p as the set of all $(y_p, q) \in x$ such that $(y, p) \in x$ and q is compatible with p. Notice that this is a definition by transfinite recursion: we can suppose when defining x_p that we have already defined y_p for all y of rank less than the rank of x, and moreover that y_p is of rank

less than or equal to the rank of y, so we can establish by an induction parallel to the recursive definition that the rank of x_p is less than or equal to the rank of x.

Note about restricted names: Notice that if we restrict names as in the note above, the recursive definition will be the same: the only incompatible things removed will be actual elements of x; elements of elements of x which are not removed will themselves not be removed because their associated conditions will be compatible with p. The entire development can proceed as below, and the recursion in the definition of x_p is removed.

weak membership and its forcing sets: $[`x'\in_0 y]$ is the smallest truth value which contains each p such that for some $q \leq p$ and x' such that $x'_p = x_p$ (notice that this is an appeal to the literal equality of reduced names in the ctm, not a recursive appeal to the membership on elements of the domain for which we define forcing sets below), we have $(x', q) \in y$. We will see below that $\in_0$ is a relation between names (as weak elements) and elements of our domain (as "weak sets"): it does not respect the membership relation on objects of the domain on the left, though it does on the right, as we will see in the next clause: this is why we put the letter on the left in quotes.

Note on the situation with restricted names: If we use restricted names, this simplifies to "$[`x'\in_0 y]$ is the smallest truth value which contains each p such that we have $(x_p, p) \in y$", and in fact the restrictions ensure that it simply is the collection of conditions p such that $(x_p, p) \in y$.

equality and its forcing sets: $[x = y]$ is defined as $[(\forall z \in N : `z'\in_0 x \leftrightarrow `z'\in_0 y)]$. For p to be in this truth value set means that for every $q \geq p$, x and y have the same elements in the sense of $\in_0$ under condition q. This definition makes it clear that we did not have to put y in quotes in the definition of $\in_0$.

Note on the situation with restricted names: With restricted names, we have a much simplified situation. If $p \in [x = y]$ we have for any condition $q \geq p$ that $(z_q, q) \in x \leftrightarrow (z_q, q) \in y$ (because z forces that z is weakly in x under condition q iff it forces that z is weakly in y) but then because of the restrictions on names we see that x_p and y_p then have exactly the same elements in the normal sense, so $x_p = y_p$.

sethood and true membership and their forcing sets: The forcing set $[\mathtt{set}(x)]$ is defined as $[(\forall zw \in N.z = w \rightarrow \text{`}z\text{'}\in_0 x \leftrightarrow \text{`}w\text{'}\in_0 x)]$. The forcing set $[x \in y]$ is then defined as $[\text{`}x\text{'}\in_0 y \wedge \mathtt{set}(y)]$.

It is possible for there to be names x, y, z and a condition p such that $p \in [y = z]$, $p \in [\text{`}y\text{'}\in_0 x]$ and $p \in [\neg\text{`}z\text{'}\in_0 x] = [\text{`}z\text{'}\notin_0 x]$. We will then have that $p \in [\neg\mathtt{set}(x)]$, because x contains one name for the element of the domain which p says is named by both y and z, and does not contain another name for the same object, under the given condition. When an object x is forced by condition p to be a set, this means that every condition $q \geq p$ which forces an equation $[y = z]$ forces $[\text{`}y\text{'}\in_0 x]$ iff it forces $[\text{`}z\text{'}\in_0 x]$.

Note on the situation with restricted names: This distinction collapses. $p \in [\mathtt{set}(x)]$ always holds, and $[x \in y]$ is the same set as $[\text{`}x\text{'}\in_0 y]$.

name closures: For every name x, we can construct a name x^* (called its "name closure") such that the conditions $[\mathtt{set}(x^*)]$ and $[(\forall y \in N :\text{`}y\text{'}\in_0 x^* \leftrightarrow (\exists z \in N.y = z \wedge \text{`}z\text{'}\in_0 x)]$ both are simply the set P. The idea is that x^* is obtained by fattening up x to obtain a name which satisfies the closure condition defining sets, that names for the same object are always either both included or both excluded once they are forced to be names of the same object and either of them is forced into or out of the set. The exact way that this is achieved is that for each condition p which forces $[y = z]$ and for which there is a condition $q \leq p$ such that $(y, q) \in x$, we add (z_p, p) to x^*.

Please note that the notation x^* for the name closure of x is only used in the preceding and following paragraphs: stars are used for various different purposes in this chapter.

We discuss the reasons why this works. First of all, we note a useful fact about $p \in [y = z]$. Any u which is forced to be a weak element of y by p or a stronger condition q is witnessed as such by an element (u', r) of y where r is compatible with p (resp. q), and $u'_p = u_p$. Now u must also be forced to be a weak element of z by the same condition, by an element (u'', r') of z where r' is compatible with p (resp. q) and $u''_p = u_p$. It follows from these considerations that y_p and z_p have the same first projections of elements: for every $(u_p, r) \in y_p$, there must be a $(u_p, r') \in z_p$ (possibly with a different condition) and vice versa. This

means that y_p and z_p, though they may not be exactly the same sets, are of the same rank, and implies that the fattening process which builds x^* does not increase rank, and so must succeed in defining a set. Further, it should be clear that we do not add weak elements to x^* which are not equal to some weak element of x: any "new" weak element z' of x^* will have z'_p equal to one of the z_p's for which (z_p, p) was added by the fattening process (driven by the fact that $p \in [y = z]$ for some y) and then $z' = y$ will also be forced by p (for the general reason that $y = z$ is forced by p iff $y_p = z_p$ is forced by p).

Note on the situation with restricted names: Name closures are not needed. All names are names for sets if restricted names are used.

As noted above, this membership relation respects equality on names. We now of course have weak extensionality: there are distinct atoms with no elements.

We will eventually throw away the atoms.

Definition (filter in P, dense set in P): Now we are going to do black magic due to the fact that we are in a countable model. A *filter* is a subset F of P which represents an effort to decide what is actually true. The defining conditions of a filter F are that if p belongs to F, and $q \leq p$ (q represents less information than p) it follows that q belongs to F as well. If two elements p and q both belong to F, they are compatible: there is r such that $r \geq p$, $r \geq q$, and $r \in F$. Now comes the black magic. We call a subset D of P *dense* iff for every $p \in P$ there is $q \geq p$ which belongs to D. A dense set of conditions is a set of conditions which is somehow impossible to avoid. We have already provided that a dense truth value is actually true.

Definition and construction of a generic filter in P: We construct a *generic filter* G, which is defined as a filter which meets every dense subset of P. This will not be an object in our ctm. We can build it, on the outside, by listing all the dense sets in a list D_i, choosing a condition p_0, and then at each step choosing p_{i+1} so that it is greater than p_i and included in D_i (since there is an element of D_i greater than any element of P). Further, of course, anything $\leq$ an element of G will belong to G. Notice that if two things belong to G, they belong because they are $\leq$ a q_i and a q_j, and both will be $\leq$ the larger-indexed of these two.

G is easily seen not to be in the ctm, because $P - G$ is dense (because we assumed the splitting property), and so cannot have been one of the D_i's.

We now use G to collapse each condition in P in effect to a truth-value, thus collapsing names to sets (with some care).

G makes our logic sensible: We claim that for every proposition ϕ, either there is $p \in G$ which forces ϕ or there is $p \in G$ which forces $\neg\phi$. The reason for this is that the set of $p \in P$ which either force ϕ or force ϕ is dense in P, and so must contain an element of G. It is also the case that we cannot have a $p \in G$ which forces ϕ and a $q \in G$ which forces $\neg\phi$, as any two conditions in G must be compatible. We refer to this as the decision property: G must in a certain sense settle the truth value of each formula.

We deal further with the logic of the forcing model. As we noted above, for every formula ϕ there is either $p \in G$ which forces ϕ or $p \in G$ which forces $\neg\phi$: the collection of conditions which either force ϕ or force $\neg\phi$ is not the set of all conditions, but it is dense, so it contains some $p \in G$. If we define $G \vdash \phi$ as $(\exists p \in G : p \in [\phi])$, we see that $G \vdash \neg\phi$ is equivalent to $\neg(G \vdash \phi)$. It should be clear that $G \vdash (\phi \wedge \psi)$ iff $(G \vdash \phi) \wedge (G \vdash \psi)$ and that $G \vdash (\forall x \in N.\phi[x])$ iff $(\forall x \in N : G \vdash \phi[x])$. We will verify below that what G says about atomic sentences comports with what happens in our model and so that G's logic agrees exactly with ours. It is particularly worth noting, though it does follow from what we have already said, that $G \vdash \phi \vee \psi$ does imply that either $G \vdash \phi$ or $G \vdash \psi$ holds: if some $p \in G$ forces $\phi \vee \psi$, it does not necessarily force one of the disjuncts, but the set of $q \geq p$ which force one of the two disjouncts is dense, and so includes an element of G. Similarly, if $G \vdash (\exists x.\phi[x])$, so there is $p \in G$ which forces $\phi[x]$, the set of $q \geq p$ for which there is a name t such that $G \vdash \phi[t]$ is dense above p, though it may not contain p, so it does contain some $q \in G$, so $G \vdash \phi[t]$ for some particular name t.

Definition (equivalence of names): Define an equivalence relation $\sim_G$ on names by $x \sim_G y \leftrightarrow (\exists p \in G : x_p = y_p)$. This is a modified version of the relation of literal identity between names, taking advantage of the generic filter to know which information is irrelevant and can be thrown away by reducing names.

Definition (elements of the first approximation to our model): Define $\underline{x}$, for any name x, as the collection of all names z such that for some $(y, p) \in x$, with $p \in G$, we have $z \sim_G y$, and moreover z is of minimal rank in the collection of names $\{w \in N \mid w \sim_G y\}$ (that is, z belongs to the lowest indexed set N_α with nonempty intersection with this set).

weak membership in model elements defined and related to forcing: We define $z \in_0 \underline{x}$ as $(\exists y \in \underline{x} : z \sim_G y)$. We claim that $z \in_0 \underline{x}$ iff there is $p \in G$ such that $p \in [\text{`}z\text{'}\in_0 x]$. Suppose $z \in_0 \underline{x}$. Then there is $u$ in $\underline{x}$ such that $z \sim_G u$, so there is $(y, q) \in x$ with $q \in G$ and $y_p = u_p = z_p$ for some $p \geq q$ also in $G$, and this establishes $p \in [\text{`}z\text{'}\in_0 x]$. Now suppose that for some $p \in G$, $p \in [\text{`}z\text{'}\in_0 x]$. This means that there is $(y, q) \in x$, with $q \leq p$, such that $y_p = z_p$. Now $q \in G$ because G is a filter, and we see that $y \sim_G z$, from which it follows that both y and z stand in the relation $\sim_G$ to some element u of minimal rank in their common equivalence class, which belongs to $\underline{x}$, so we have established that $z \in_0 \underline{x}$.

Note that we have shown that $G \vdash \text{`}z\text{'}\in_0 x$ iff $z \in_0 \underline{x}$.

equality of model elements related to forcing: Now we claim that $\underline{x} = \underline{y}$ iff for some $p \in G$, $p \in [x = y]$.

1. Suppose $\underline{x} = \underline{y}$. We need to show that for some $p \in G$, $p \in [x = y]$. It appears easier to prove the contrapositive. Suppose that for no $p \in G$ do we have $p \in [x = y]$. The set of conditions q such that $q \in [x = y]$ or $q \in \neg[x = y]$ is dense in P, so there must be $p \in G$ such that $p \in [\neg x = y]$ (decision property). This means that p forces $(\exists z : (\text{`}z\text{'}\in_0 x \wedge \neg\text{`}z\text{'}\in_0 y) \vee (\text{`}z\text{'}\in_0 y \wedge \neg\text{`}z\text{'}\in_0 x))$, so for some stronger $q \in G$ we have a specific name z such that either q forces $\text{`}z\text{'}\in_0 x$ and $q$ forces $\neg\text{`}z\text{'}\in_0 y$, or vice versa, which implies $z \in \underline{x} \wedge z \notin \underline{y}$, or vice versa, and in either case $\underline{x} \neq \underline{y}$ as desired.
2. Suppose that some $p \in G$ forces $x = y$. We aim to show that $\underline{x} = \underline{y}$. Suppose that $z \in_0 \underline{x}$. It follows that for some $q \in G$, q forces $z \in_0 x$. A condition r stronger than both p and q will force $z \in_0 x$ and $x = y$, or equivalently $z \in_0 x \leftrightarrow z \in_0 y$, so in fact it also forces $z \in_0 y$ so $z \in \underline{y}$. The argument is symmetrical so $\underline{x} = \underline{y}$.

Note that we have shown that $G \vdash x = y$ iff $\underline{x} = \underline{y}$.

sethood and membership proper defined for model elements: We now define $\mathtt{set}(\underline{x})$ as holding iff $(\forall zw : \underline{z} = \underline{w} \rightarrow (z \in_0 x \leftrightarrow w \in_0 x)$, and define $\underline{x} \in_1 \underline{y}$ as holding iff $x \in_0 \underline{y} \wedge \mathtt{set}(\underline{y})$. Arguments precisely similar to ones we have already given show that $\mathtt{set}(\underline{x})$ holds exactly if $\mathtt{set}(x)$ is forced by some $p \in G$, and $\underline{x} \in_1 \underline{y}$ holds iff there is p in G which forces $x \in y$.

We are claiming that $G \vdash \mathtt{set}(x)$ iff $\mathtt{set}(\underline{x})$ and $G \vdash x \in y$ iff $\underline{x} \in_1 \underline{y}$, completing the verification that G assigns values to atomic sentences in a way which comports with our model.

with restricted names... $\in_1$ will coincide with $\in_0$ and there will be no atoms. $G \vdash \mathtt{set}(x)$ will always hold.

membership of model elements is extensional: That the relation $\in_1$ is (weakly) extensional should be clear. We have shown above that the model elements $\underline{x}$ are equal iff they have the same names as elements in the weak sense; the same is true if we restrict our attention to the model elements treated as sets, and their extensions under $\in_1$ determine their extensions under $\in_0$, so sets with the same associated extension of model elements under $\in_1$ have the same associated extension of names under $\in_0$, and so are equal. Model elements $\underline{x}$ of which $\mathtt{set}(\underline{x})$ does not hold have no elements in the sense of $\in_1$: they are treated as atoms.

with restricted names... Membership of model elements will be strongly extensional.

membership of model elements is well-founded: Now we need to show that $\in_1$ is a well-founded relation. For any model element $\underline{x}$, we define the rank of x as the smallest rank of a name y such that $\underline{x} = \underline{y}$. Now suppose that $\underline{z} \in_1 \underline{x}$, where x has the same rank as $\underline{x}$. This implies that $\underline{z} \in_0 \underline{x}$, from which we can conclude that for some $p \in G$ we have that p forces 'z'$\in_0 x$, from which we can conclude that there is u with $u_p = z_p$ (for which p forces $u = z$) and $q \leq p$ such that $(u, q) \in z$. Now $\underline{u} = \underline{z}$ and the rank of u is less than the rank of x, so the rank of $\underline{z}$ is less than the rank of $\underline{x}$, and we conclude that $\in_1$ is well-founded.

the forcing model proper constructed by a modified Mostowski collapse: Thus we can carry out a modified Mostowski collapse sending each model element $\underline{x}$ to a new model element $\underline{x}^*$ with $\underline{x}^* = (\underline{x}, \omega_1)$ for each

atom $\underline{x}$ (that is the real uncountable ω_1, not in our ctm), and $\underline{x}^*$ defined (by transfinite recursion on rank) as $\{\underline{y}^* \mid \underline{y} \in_1 \underline{x}\}$. when $\mathtt{set}(\underline{x})$. Notice that we will have $\underline{x}^* = \underline{y}^*$ iff $\underline{x} = \underline{y}$, by an induction parallel to the recursion in the definition. The use of the real ω_1 is a technical device to prevent the accidental construction of the set implementing an atom as the set implementing some set as well.

In what follows, our model elements are the objects $\underline{x}^*$, but we will refer to an object of our model $X = \underline{x}^*$ as the object X with name x (and of course many other names as well). We refer to X as the "collapse" of x. We then free up the star to mean something else, as we will no longer use it to refer to the product of the Mostowski collapse carried out here.

with restricted names... There are no atoms so the Mostowski collapse is unmodified, and the curious case with the true ω_1 involving atoms does not occur.

We then hope to restrict our attention entirely to the pure sets, those whose transitive closures contain no atoms. **with restricted names**... there is no need to do this.

As we will see, this process of collapse will sometimes produce sets which were not in our original ctm (certainly atoms are not in the ctm, nor is any set with an atom in its transitive closure: the question is what pure sets are present). Copies of all sets in the original ctm will be present: if we have names $\hat{y}$ for all elements of a set x in the original ctm, define $\hat{x}$ as the name closure of $\{\hat{y} \mid y \in x\} \times P$, and $\hat{x}$ will name x.

We want to verify that all axioms of ZFC actually hold in the new structure (with the qualification that extensionality is weakened).

Extensionality: If x and y have elements and have the same elements in the new structure, they are certainly equal. I have a feeling there is more to be said about this.

Pairing: If x and y have names x^* and y^*, take the name $(x^* \times P) \cup (y^* \times P)$, close it up so it names a set, and it will collapse to $\{x, y\}$.

Union: If x has a name x^*, define a name z^* for the union of the object named by x^* as follows: for each (y^*, p) in x^* and each $(u^*, q) \in y^*$, z^* must contain each (u^*, r) such that $r \leq p$ and $r \leq q$; z^* is the result of

applying name closure to the smallest relation compatible with these conditions. The collapse of z^* will be the union of the collapse of x^*.

Power Set: Let x^* be a name for x. Construct all possible names y^* relating each element of the domain of x^* to a possibly proper subset of the upward closure of the set of elements of P to which x^* relates it. Let Y be the set of such names. Take the name $Y \times P$ and close it up if necessary so that it names a set. This will name the power set of x.

Separation: To build $\{x \in A \mid \phi(x)\}$, relate each name x^* in the domain of A^* to the intersection of $[x^* \in A^*]$ and $[\phi[x^*]]$. Close the result so that it names a set.

Infinity: $\hat{\omega}$.

Choice: For any name x^*, build a name for a map from an ordinal α or an initial segment thereof to the elements of x: well-order the domain of x and let α be the order type of this well-ordering. Build a name for a function by building a name for a set of ordered pairs, sending each y^* to $\beta < \alpha$ (using the name $\hat{\beta}$ for each such ordinal) under any condition under which each preceding domain element of x has either been shown not to be in x or already mapped to some $\gamma < \beta$. Close the name computed so that it is the name of a set.

Replacement: To compute the image of a set with name A^* under a putatively functional formula $\phi[x, y]$, take as your domain all names y^* of minimal set theoretic rank such that for some condition p, p says that $\phi[x, y]$ is functional and p says that $\phi[x^*, y^*]$, for some x^* such that $p \in [x \in A]$, and relate each y^* to each element of the smallest truth value set containing all such conditions. There is some trickiness about seeing that Replacement in the original ctm ensures that the name is actually a set: the trick is contained in the requirement that we choose names of minimal rank. Close the name up if necessary to ensure that it names a set.

The new structure is still countable, because it is no larger than the collection of names, which is a subset of the original ctm. For every object x in the original ctm, there is a name $\hat{x}$ defined as the name closure of $\{\hat{y} \mid y \in x\} \times P$, which in fact names the original set x. There is at least one new object in the structure, namely the collapse of the name closure of

$\{(\hat{p}, p) \mid p \in P\}$, which contains p just in case $p \in G$. This is precisely the set G, which we know was not in the original ctm (it cannot be, because $P \setminus G$ is dense in P by the splitting property, and G meets every dense subset of P in the ctm).

Finally, we can expunge the atoms because we can define what it means to be an atom (not equal to $\emptyset$ and having no elements) and then we can define what it means to be a pure set (having no atoms in one's transitive closure), and restrict our domain to pure sets. The domain of pure sets will still satisfy all the axioms, and thus will be a countable transitive model of ZFC.

Exercises

1. Please carefully read the section on forcing and communicate with me about any typos or errors you find and any points that confuse you.

2. Prove that for any formulas ϕ and ψ and condition p, p forces $\phi \to \psi$ if and only if $(\forall q \geq p : q \in [\phi] \to q \in [\psi]\}$.

3. Draw a picture of conditions arranged in a finite diagram (with the relation $\leq$ between conditions indicated) with indications of where propositions are forced satisfying the following perhaps unexpected conditions. These may *all* be things I actually sketched on the board.

   (a) A condition p (with incompatible conditions q and r pictured above it) forces $\phi \vee \psi$ without forcing either ϕ or ψ (next to each condition in your diagram, indicate which atomic statements it forces).

   (b) A condition p (with incompatible conditions q and r pictured above it) forces $(\exists x \in N : \phi[x])$ while there is no name y such that p forces $\phi[y]$ (again, label each condition with the set of atomic statements that it forces: you may leave the predicate $\phi[x]$ completely abstract and use arbitrary names like a, b.

   (c) In this part you have to reason about names. Give a diagram of conditions and actual names x, y (presented as subsets of $N \times P$: the names in the domains of x and y may be letters a, b) such that incompatible conditions q and r force $x = y$ and $x \neq y$ respectively.

(d) Give an example of a name x and a condition p such that p forces $a \in x$ (a can just be a letter) but neither p nor any condition weaker than p is in the range of x (there is no pair of the form (u, p) or even of the form (u, q) with $q \leq p$ belonging to x).

4. Prove that if $x.y \in N$, and $x_p = y_p$, and $q \geq p$, then $x_q = y_q$. Notice that this proof will involve transfinite induction: you prove this assertion for x and y with the maximum of the ranks of x and y being α under the assumption that it is known to be true for all u, v with the ranks of both u and v less than α. If you get after me, I'll try to prove some result by a similar induction in class.

   Partial Solution: Since I had the definition of x_p wrong in the notes, I'll give a strong hint on this one. To begin with, we might want to make sure that the notation x_p is well-defined. We do this by defining a notation x_p^α decorated with an ordinal rank. If $x \in N_\alpha$, we define x_p^α as the set of all (y_p^β, q) such that $\beta < \alpha$, $y \in N_\beta$, $(y, q) \in x$, and q is compatible with p. This is more clearly a definition by transfinite recursion: notice that for any $(y, q) \in x$, y has to belong to an N_β with $\beta < \alpha$, so we have already defined y_p^β.

   We argue by transfinite induction on α that if $x \in N_\alpha$, then $x_p^\gamma = x_p^\alpha$ for all $\gamma > \alpha$: Suppose (z, q) belongs to x_p^α, Then q is compatible with p, and for some y, $z = y_p^\beta$ with $\beta < \alpha$ and $y \in N_\beta$ and $(y, q) \in x$. Note, though, that $\beta < \alpha$ implies $\beta < \gamma$ as well, so $(z, q) \in x_p^\gamma$ on the same evidence. The converse is slightly harder. Suppose (z, q) belongs to x_p^γ. It follows that q is compatible with p and for some $\beta < \gamma$ and $y \in N_\beta$ we have $z = y_p^\beta$ and $(y, q) \in x$. Now, because $(y, q) \in x$, we actually have $y \in N_{\beta'}$ for some $\beta' < \alpha$ (not merely $< \gamma$). By inductive hypothesis we have $y_p^{\beta'} = z = y_p^\beta$, and so we also have the evidence required that $(z, q) \in x_p^\alpha$.

   I'll lecture the full solution on Wednesday (and put it here in the notes). For the moment, I'll give you a hint. You now have everything you need to prove by transfinite induction that if $q \geq p$, then $(x_p)_q = x_q$. Prove this by transfinite induction on ranks of names, and the result claimed above follows immediately.

5. Prove that if $x_p = y_p$ for any condition p, it follows that $p \in [x = y]$, and further that if $p \in G$ in addition it follows that $\underline{x} = \underline{y}$. I believe that the result of the previous problem will be used in this argument.

6. (challenge problem, I have no idea if this is reasonable): Suppose that a condition p forces the condition that x^* is the name of a partition. Try to construct a name y^* such that the same condition p will force y^* to be a choice set for x^*.

3.10.1 Independence of CH

Now build the classic example. We construct a model which believes that there is a set of real numbers of size $\aleph_2$, demonstrating independence of the Continuum Hypothesis.

Our partial order will be the inclusion relation (not the inverse inclusion relation, as is usually stated, because we use the converse order from usual presentations; this is purely a technicality) on the set of finite functions from subsets of $\omega_2 \times \omega$ to $\{0, 1\}$.

The union of a generic filter G on this set will be a function from $\hat{\omega_2} \times \omega$ to $\{0, 1\}$ (where $\hat{\omega_2}$ here means the fake ω_2 of the countable transitive model, which is actually some countable ordinal). To see this, observe that each element of the generic filter will be a finite function from a subset of $\hat{\omega_2} \times \omega$ to $\{0, 1\}$, and for any $\alpha \in \hat{\omega_2}$ (the fake ω_2) the set D of conditions p such that (α, m) is in the domain of p is dense (because any condition at all can be extended with $((\alpha, m), 0)$ or $((\alpha, m), 1)$ as desired, so G meets D, and of course all elements of G must agree on which of 0 or 1 is supplied as a value, because any two elements of G must be compatible).

For each $\alpha \in \hat{\omega_2}$, there will be a subset $r_\alpha = \{m \in \omega \mid ((\alpha, m), 1) \in \bigcup G\}$, and the forcing model will contain a function sending each $\alpha \in \hat{\omega_2}$ to r_α: a name for this is easy to construct – put a pair (α, m) into the name under condition p (that is, add a pair (α, p) as an element of the name) just in case $((\alpha, m), 1) \in p$.

All the r_α's are different: if p is any condition and $\alpha \neq \beta$ are elements of $\hat{\omega_2}$, p can always be extended with $((\alpha, n), 0)$ and $((\beta, n), 1)$ for a large enough n, so the set of conditions under which the forcing model must put some number n into r_α and exclude it from r_β is dense, and so such a condition must belong to G.

Thus the forcing model contains a bijection from $\hat{\omega_2}$ to the natural numbers, which would seem to imply that we were done. However, there is a little more work to do. The problem is that it is not obvious that $\hat{\omega_2}$ is actually ω_2 in the opinion of the forcing model (as it is in the option of the original ctm).

We briefly describe a different forcing model to show that there is a real issue. Use the partial order of inclusion on finite injective maps from subsets of ω to subsets of $\hat{\omega_2}$. Let H be a generic filter in this partial order. The union of H will be a bijection from ω to $\hat{\omega_2}$, and so the resulting forcing model will think (correctly, unlike the original ctm) that $\hat{\omega_2}$ is a countable ordinal!

We say that a partial order $\leq$ on P (in the original ctm) satisfies the ccc (countable chain condition: the name is traditional, but an error: it is really the countable antichain condition) iff there is no uncountable set of mutually incompatible elements in the field of $\leq$. We show that if forcing under $\leq$ creates a bijection between distinct cardinals in the original ctm, then $\leq$ fails to satisfy the ccc in the original ctm.

Let $\kappa < \lambda$ be infinite cardinals, and let f be a name in the forcing model produced from P for a bijection from κ to λ. λ is uncountable in the original ctm. For each $\alpha < \lambda$, there must be a condition p which forces for some $\beta < \kappa$ the assertion $f(\beta) = \alpha$. Now we use the Pigeonhole Principle. For each α in λ there are one or more β's in κ which work. Because λ is greater than κ, there must be a specific β which works for λ distinct α's. But then there are conditions p_α for λ distinct α's such that for one and the same β, p_α forces $f(\beta) = \alpha$, and these p_α's make up a pairwise incompatible collection of conditions, violating the ccc.

It follows that we can show that $\hat{\omega_2}$ is the ω_2 of our original forcing model (because the forcing model will have the same cardinals as the original ctm) if we can show that the partial order we started with has the ccc.

Suppose that we have an uncountable set of incompatible conditions in our original partial order and argue to a contradiction.

Choose a single condition p_0 (partial function from a subset of $\omega_2 \times \omega$ to $\{0, 1\}$) in the uncountable mutually incompatible set of conditions. Each pair of conditions in this set has associated with it the subset of $\omega_2 \times \omega$ on which both conditions have values and the values disagree.

For the single condition p_0 we have chosen, there are only finitely many possible values for this distinguished subset in relation to any other condition, and there are countably many values for the size of the other condition. So there is an uncountable collection B_0 of conditions in the uncountable mutually incompatible set which have the same set of disagreement in relation to the single condition p_0 chosen initially and which are all the same size.

So we now have a mutually incompatible uncountable set of conditions all of which agree on a certain finite subset of their domains, and all of which

are the same size as sets. We can choose a single element of this mutually incompatible set and repeat the process: choose an element p_1 of B_0 and construct a subset B_1 of B_0 all elements of which are the same size (in fact the common size of elements of B_0) and all of which disagree with p_1 on the same nonempty finite set. However, this will fail after finitely many steps, because when we construct each p_n, we discover a new set B_n all of whose elements disagree with each of the p_i's constructed so far and agree with other on a finite subset of their domains, larger at each step, and all elements of each B_n are of the same size (the common size of all the elements of B_0).

Thus our partial order satisfies the ccc, and our forcing model has at least ω_2 distinct subsets of the natural numbers, and so has ω_2 distinct real numbers, so CH fails there.

3.11 Independence of Choice

In this section, we will explicitly show that the axiom of choice is independent of ZFC if atoms are allowed.

The modified set theory ZFCA we work in has Extensionality weakened to allow atoms. Further, we assert that there is a set $\mathbb{A}$ of all atoms. The other axioms are as before (including choice).

It is straightforward to interpret ZFCA with the set of atoms of any desired size in ZFC. Choose a set A. We will redefine the membership relation $\in$ to give a new relation $\in'$ under which the axioms of ZFCA will hold, and all and only the elements of $A \times \{0\}$ become atoms. The definition is $x \in' y$ iff either $y \notin A \times \mathbb{N}$ and $x \in y$ or $y = (a, n+1)$ for some $a \in A$ and $n \in \mathbb{N}$ and $x \in (a, n)$. We leave it to the reader to work out the details.

In ZFCA, there is a cumulative hierarchy as there is in ZFC:

We give a definition by transfinite recursion of the hierarchy in ZFCA:

1. $V_0 = \mathbb{A}$.

2. $V_{\alpha+1} = V_\alpha \cup \mathcal{P}(V_\alpha)$. (this definition preserves the idea that the levels of the hierarchy are cumulative. If we simply took power sets, we would for example first construct sets with some elements sets and some elements atoms at level ω, which would be odd.)

3. $V_\lambda = \bigcup_{\beta<\lambda} V_\beta$ for λ limit.

We are interested in bijections from $\mathbb{A}$ to $\mathbb{A}$, which we refer to as permutations of the atoms. For each permutation π, we indicate how to extend the definition of the notation $\pi(A)$ to sets A. We do this by defining a sequence π^α of functions by transfinite recursion:

We define π^0 as the permutation π itself, considered as a map from V_0 to V_0.

When we have defined π^β for each $\beta < \alpha$, a permutation of V_β, and moreover we have that each pair π^β, π^γ for $\beta < \gamma < \alpha$ agree on the intersection V_β of their domains, we indicate how to define V_α.

If α is a successor $\delta + 1$. we define $\pi^\alpha(A)$ for each $A \in V_\alpha$ as $\pi^\delta\text{“}A$ on $\mathcal{P}(V_\delta) - V_\delta$ and as $\pi^\delta(A)$ for $A \in V_\delta$.

If A is in the intersection of V_δ and $\mathcal{P}(V_\delta)$, A must belong to some $V_{\epsilon+1} - V_\epsilon$ where $\epsilon < \delta$, and $\pi^{\epsilon+1}(A) = \pi^\epsilon\text{“}A$ agrees with $\pi^\delta\text{“}A$ and with $\pi^\delta(A)$, in both

cases by the inductive hypothesis on agreement of functions with index less than α.

It is then clear, since π^δ is a permutation of V_δ, that π^α is a permutation of $V_\alpha = V_{\delta+1}$.

We also need to prove that π^α agrees with each π^β for $\beta < \alpha$. Suppose this were not the case. Then there would be a minimal β such that π^β disagreed with π^α at some x, and for this value of β, a minimal γ such that such an x could be found in $V_{\gamma+1} - V_\gamma$ (it is quite clear that such an x will not be an atom). But it then follows that π^α (and π_δ) agrees with π_γ and π^β on all the elements of x, whence $\pi^\alpha(x) = \pi^\delta\text{“}x = \pi^\gamma\text{“}(x) = \pi^{\gamma+1}(x) = \pi^\beta(x)$ by the way the functions are defined and the inductively hypotheses about agreement.

Now if α is limit, we define $\pi^\alpha(x)$ for each $x \in V_\alpha$ as the common value of all $\pi^\beta(x)$'s for $x \in V_\beta$ (common by inductive hypothesis). This defines a permutation of V_α because we know by inductive hypothesis that each of the π_β's included in it is a permutation of V_β. It agrees with all lower indexed maps directly by the way it is defined.

We can then define $\pi(A)$ (admittedly an abuse of notation) as $\pi^\alpha(A)$ for any α such that $A \in V_\alpha$. We have thus converted the set of permutations of the atoms into a set-sized collection of class permutations acting on the entire universe.

Observe that for any x and y, $x = y \leftrightarrow \pi(x) = \pi(y)$ and $x \in y \leftrightarrow \pi(x) \in \pi(y) = \pi\text{“}y$. Further observe that $(\forall x : \phi[x])$ is equivalent to $(\forall x : \phi(\pi(x)))$ (and ditto for existential quantifiers) because π acts as a permutation on the entire universe.

From this it follows for any formula $\phi[x, a_1, \ldots, a_n]$ with all free variables listed in the vector that $\phi[x, a_1, \ldots, a_n] \leftrightarrow \phi[\pi(x), \pi(a_1), \ldots, \pi(a_n)]$. From this it follows that if $\{x \mid \phi[x, a_1, \ldots, a_n]\}$ exists, then $\pi(\{x \mid \phi[x, a_1, \ldots, a_n]\}) = \{\pi(x) \mid \phi[x, a_1, \ldots, a_n]\} = \{\pi(x) \mid \phi[\pi(x), \pi(a_1), \ldots, \pi(a_n)]\} = \{x \mid \phi[x, \pi(a_1), \ldots, \pi(a_n)]\}$ (the last move exploiting the fact that π acts as a permutation of the entire universe).

We define a subclass of the universe relative to any group G of permutations of the atoms. We say that an object x has support S, a finite set of atoms, iff every permutation $\pi \in G$ such that for each $s \in S$, $\pi(s) = s$, also satisfies $\pi(x) = x$. We claim that for any group G of permutations, the class of sets (and atoms) with support satisfies all the axioms of ZFCA except possibly Choice.

That Extensionality holds is evident. The reasons why all the other axioms hold are exactly the same. Each such axiom asserts the existence

of a more or less complicated set $\{x \mid \phi[x, a_1, \ldots, a_n]\}$, with parameters $a_1, \ldots, a_n$. Each parameter a_i (being taken from our subclass) has a support S_i. The union of the sets S_i is finite, and any permutation π in G which fixes each element of $\bigcup_{1 \leq i \leq n} S_i$ will fix each a_i and so by the formula $\pi(\{x \mid \phi[x, a_1, \ldots, a_n]\}) = \{x \mid \phi[x, \pi(a_1), \ldots, \pi(a_n)]\}$ proved above will fix $\{x \mid \phi[x, a_1, \ldots, a_n]\}$.

Now we show that choice does not hold in some specific models of this kind. Let G be the entire group of permutations of the atoms. Let $\mathbb{A}$ be infinite. A set $A \subseteq \mathbb{A}$ will have finite support S in the set of all permutations of the atoms only if it is finite or if $\mathbb{A} - A$ is finite. So this model of ZFA contains a set $\mathbb{A}$ which it believes to be infinite and to have no subset A such that both A and $\mathbb{A}$ are infinite. It is straightforward to prove in ZFC or ZFCA that an infinite set must have a subset of size ω, which in turn has two disjoint infinite subsets, so choice fails in this model.

We introduce another model to implement Russell's famous example of infinitely many pairs of socks. Let the set of atoms $\mathbb{A}$ support a partition P into infinitely many two element sets (the atoms are socks and the elements of P are pairs). Let G be the set of all permutations of $\mathbb{A}$ whose action fixes each element of P (each permutation may fix each sock or exchange it with its mate). We claim that no choice set C for P can have finite support with respect to these permutations. Suppose it did have a support S. S is a finite set, so there is a pair of socks $p = \{a, b\}$ which is disjoint from S. The permutation which sends a to b, b to a, and fixes every other sock belongs to the group G, fixes each element of S, but moves C, so S was not a support of C: this contradiction makes our point.

The pairs of socks example is nice because it actually gives us an explicit failure of the Axiom of Choice as usually written: we actually have a partition P of the atoms which is in the model (P is fixed by the action of every permutation in G, so it has support the empty set!) which does not have a choice set in the model. It is less obvious how to get a partition without a choice set in the first model where we used all permutations of the atoms (it is not going to be a partition of the set of atoms, because the set of atoms has no infinite partitions in that model!).

Unfortunately, we have not shown that Choice fails in ZFC by this argument. The problem is that the well-founded sets of ZFC (without the wiggle room afforded by atoms) are a rigid structure: we have no way to define a permutation π of the universe such that $\pi(A) = \pi``A$ for every A which is not simply the identity.

We will outline vaguely how to do this using forcing technology. In fact, we can use the exact model we used to establish the independence of Choice. The key is to use permutations of the partial order P used in our forcing to define a notion of support for *names*, and to allow only names with support in our forcing model. The partial order P used in the construction above was the inclusion order on finite subsets of $\omega_2 \times \omega$. Our group of permutations acts on the ω_2 columns of the set $\omega_2 \times \omega$ (in effect permuting the ω_2 generic subsets of ω we were constructing in the original construction, without permuting the other dimension which tells which natural numbers belong to each of the subsets of ω). A name will have support S (a finite subset of ω_2) of any permutation of the columns which fixes all the elements of S fixes the name. It is technically exciting to show that our forcing constructions go through if we only allow use of names with support in this sense. In the forcing model we end up with, we have certainly not added ω_2 reals, but we have added a large set of reals on which it is impossible to put a well-ordering (it seems quite reasonable that it would be difficult to produce a symmetric name for such an ordering, at any rate), so choice fails. It will not be the case that this set of reals is as formless as the set of atoms in our first construction of atoms: even without choice, one can show that there cannot be an infinite set of reals which cannot be partitioned into two infinite sets: being a set of reals, this collection certainly has a linear order (since the reals do!), which the set of atoms in the first model construction above cannot have. I want to say more about this...

Exercises

1. Show without the Axiom of Choice that any finite partition has a choice set. Then show in the first model above that any partition of the set of atoms has a choice set.

   A serious challenge is to present a set and a partition of that set in the first model (using all permutations of the infinite set of atoms) which does not have a choice set. It is possible to reverse engineer what the set and the partition should be from a proof of the Well-Ordering Theorem or Zorn's Lemma...

2. Show that the set of atoms in each of the two models we have presented cannot be linearly ordered. The strategy is to suppose that there is a linear order realized by a set in our model, choose a finite support of

this set (which must exist by definition of the model) then show that in fact the support...cannot be a support.

A challenge: produce a model in the same way (using a different group of permutations G) which gives a model in which the atoms are linearly ordered but not well-orderable. I wouldn't regard it as implausible that you could come up with the right group of permutations: showing that it would kill a well-ordering might be hard.

3. Show that the construction of a model of ZFCA from a model of ZFA at the beginning of the section works. Again, this is a serious challenge. See if you can verify a few axioms.

4. Prove in ZFA that if a definable class permutation of the universe π has the property $\pi(A) = \pi``A$ for every set A, then it is the identity permutation: $\pi(A) = A$ for every set A.

3.12 † Bridges from untyped set theory to typed set theory

This subsection introduces relationships between the untyped theory of sets developed above and the typed theory of sets developed previously.

3.12.1 †The intended interpretation of Zermelo set theory in set pictures; the Axiom of Rank; transitive closures and Foundation

Our intention in this section is to show how Zermelo set theory can be interpreted in subsets of the set Z of set pictures with the relation E standing in for membership, and to observe that when Zermelo set theory is implemented in this way certain additional axioms hold which make the system easier to work with.

Any sentence of the language of untyped set theory can be translated into a sentence of our type theory by replacing each occurrence of $\in$ with the relation E and bounding each quantifier in the set Z (all in some fixed type). In fact, instead of bounding it in Z, we bound it in $\mathbb{E}_\lambda$, where $\lambda \geq \omega \cdot 2$ is a limit ordinal. We assume that each rank below rank λ is complete, so we are assuming at least the existence of $\beth_\omega$.

We claim that (under the assumption that all types below λ are complete), the translations of the axioms of Zermelo set theory into the language of type theory are true, so we have a way to understand untyped set theory in terms of our type theory.

Extensionality: Sets with the same elements are the same. Zermelo allowed atoms (non-sets) in his original formulation, and we allow for that possibility in the previous presentation of his axioms, but we will assume here that all objects are sets.

Verification of Extensionality: This follows from the fact that E is a membership diagram, and so an extensional relation (and the fact that E end extends the restriction of E to any $\mathbb{E}_\lambda$; the preimage of any element of the field of the restriction under the restriction is the same as its preimage under E itself, so extensionality of E implies extensionality of the restriction.

Elementary Sets: The empty set $\emptyset$ exists. For any objects x and y, $\{x\}$ and $\{x, y\}$ are sets.

Verification of Elementary Sets: The equivalence class of the empty set diagram belongs to $\mathbb{E}_\lambda$ and has empty preimage under E, so satisfies the translation of the properties of the empty set. Let $a, b \in \mathbb{E}_\lambda$. $a \in \mathbb{E}_\alpha$ and $b \in \mathbb{E}_\beta$ for some $\alpha, \beta < \lambda$. Because λ is limit, $\max(\alpha, \beta) + 1 < \lambda$, and since $\mathbb{E}_{\max(\alpha,\beta)}$ is a complete rank, $\{a, b\}$ has an E-code in $\mathbb{E}_{\max(\alpha,\beta)+1} \subseteq \mathbb{E}_\lambda$

Separation: For any property $\phi[x]$ and set A, the set $\{x \in A \mid \phi[x]\}$ exists.

Verification of Separation: Any $A \in \mathbb{E}_\lambda$ belongs to some rank $\mathbb{E}_{\alpha+1}$ for $\alpha < \lambda$ (every element of Z first appears in a successor rank). The formula $\phi[x]$ translates to a formula $\Phi[x]$ in the language of type theory. The set $\{x \in \mathbb{E}_\alpha \mid x\, E\, A \wedge \Phi[x]\}$ exists by comprehension in type theory and has an E-code in $\mathbb{E}_{\alpha+1}$ because $\mathbb{E}_\alpha$ is a complete rank.

Power Set: For any set A, the set $\{B \mid B \subseteq A\}$ exists. The definition of $A \subseteq B$ is the usual one.

Verification of Power Set: Any $A \in \mathbb{E}_\lambda$ belongs to some rank $\mathbb{E}_{\alpha+1}$ for $\alpha < \lambda$ (every element of Z first appears in a successor rank). $\alpha + 1$ and $\alpha + 2$ are also less than λ because λ is limit. The translation of $B \subseteq A$ asserts that the E-preimage of B is a subset of the E-preimage of A. Each B whose E-preimage is a subset of the E-preimage of A also belongs to $\mathbb{E}_{\alpha+1}$ (because each element of its E-preimage belongs to $\mathbb{E}_\alpha$ and $\mathbb{E}_\alpha$ is complete), so the set of all such B has an E-code in $\mathbb{E}_{\alpha+2}$, because $\mathbb{E}_{\alpha+1}$ is complete.

Union: For any set A, the set $\bigcup A = \{x \mid (\exists y \in A.x \in y)\}$ exists.

Verification of Union: Any $A \in \mathbb{E}_\lambda$ belongs to some rank $\mathbb{E}_{\alpha+1}$ for $\alpha < \lambda$ (every element of Z first appears in a successor rank). The translation of $(\exists y \in A.x \in y)$ into the language of type theory asserts that x is in the $E|E$-preimage of A, which is a subset of $\mathbb{E}_\alpha$, so has an E-code in $\mathbb{E}_{\alpha+1}$, because $\mathbb{E}_\alpha$ is complete.

Infinity: There is a set I such that $\emptyset \in I$ and $(\forall x.x \in I \to \{x\} \in I)$.

Verification of Infinity: Define a relation on the ordinals $\leq \omega$ by $x\,R\,y \leftrightarrow y = x + 1 \vee y = \omega$. The isomorphism type of this relation is the implementation of I.

Choice: Any pairwise disjoint collection of nonempty sets has a choice set.

Verification of Choice: The translation of the property "P is a pairwise disjoint collection of nonempty sets" into the language of type theory is "P is an element of $\mathbb{E}_\lambda$ such that the E-preimages of the elements of its E-preimage are nonempty and disjoint". $P \in \mathbb{E}_\lambda$ belongs to some rank $\mathbb{E}_{\alpha+1}$ for $\alpha < \lambda$. Each element of the E-preimage of an element of the E-preimage of P belongs to $\mathbb{E}_\alpha$. By the Axiom of Choice in type theory, the pairwise disjoint collection of nonempty E-preimages of the elements of the E-preimage of P has a choice set, which is a subset of $\mathbb{E}_\alpha$, so has an E-code because $\mathbb{E}_\alpha$ is a complete rank.

Furthermore, the translation of the axioms of Zermelo set theory into the theory of all set pictures expressed with type-free set picture variables are true, with a qualification, for essentially the same reasons given above. The qualification is that Separation will only work for formulas in which every quantifier is bounded in a set, because we cannot translate sentences which do not have this property from the language of type-free set picture variables back into the language of type theory. The version of Zermelo set theory with this restriction on Separation is called "bounded Zermelo set theory" or "Mac Lane set theory", the latter because Saunders Mac Lane has advocated it as a foundational system. Notice that the translation of Mac Lane set theory into the type-free theory of set pictures does *not* require the assumption that $\beth_\omega$ exists: the only axiom that requires that λ be limit in the development above is Power Set, and the verification of the translation of Power Set in the theory of all set pictures is given at the end of the section om the hierarchy of ranks of set pictures (basically, one can introduce the "power set" of any particular "set" one mentions by working in a higher type).

We state an additional axiom which holds in both the implementations of Zermelo set theory given here, but which fails to hold in some eccentric models of Zermelo set theory. This axiom expresses the idea that every element of $\mathbb{E}_\lambda$ belongs to some rank $\mathbb{E}_\alpha$.

Observation: The Kuratowski pair $\{\{x\}, \{x, y\}\}$ of two sets x and y is easily seen to be a set, and the proof that this is a pair goes much as in type

theory. We can then define relations (and in particular well-orderings) just as we did in type theory.

Definition: A *subhierarchy* is a set H which is well-ordered by inclusion and in which each successor in the inclusion order on H is the power set of its predecessor and each non-successor in the inclusion order on H is the union of all its predecessors in that order. A *rank* is a set which belongs to some subhierarchy.

Theorem: Of any two distinct subhierarchies, one is an initial segment of the other in the inclusion order. So all ranks are well-ordered by inclusion.

Axiom of Rank: Every set is a subset of some rank.

Verification of the Axiom of Rank: Each $A \in \mathbb{E}_\lambda$ belongs to some $\mathbb{E}_\alpha$, $\alpha < \lambda$. Each $\mathbb{E}_\alpha$ has an E-code, which we will call V_α, because it is a complete rank. For any β, $\{V_\alpha \mid \alpha < \beta\}$ has an E-code, which we will call H_α, because it is a subset of $\mathbb{E}_{\beta+1}$. It is straightforward to verify that H_α satisfies the stated properties for a subhierarchy (translated into the language of type theory), whence we have the translation of "V_α is a rank", and "$A \subseteq V_\alpha$", so the translation of "A is a subset of some rank" holds.

The Axiom of Rank has many useful consequences. We give two of them here.

Definition: We say that a set A is *transitive* iff $(\forall x \in A.(\forall y \in x.y \in A))$. It is worth noting that a set is transitive (in our interpretation in type theory) iff any set diagram belonging to the set picture implementing A is a transitive relation.

Theorem: Every set is included in a transitive set.

Proof: It is straightforward to prove by transfinite induction along the inclusion order that all ranks are transitive. By the Axiom of Rank every set is included in a rank.

Definition: For any set A, we define r_A as the minimal rank in the inclusion order including A as a subset. We define $\mathtt{TC}(A)$, the *transitive closure* of A, as $\{x \in r_A \mid$ every transitive set including A includes $x\}$. This

exists by Separation and is the minimal transitive set in the inclusion order which includes A as a subset. $\mathtt{TC}(\{A\})$, which also contains A as an element, will sometimes be of more interest.

Observation: That sets have transitive closures is not provable in Zermelo set theory as originally formulated. The usual proof in *ZFC* requires the very powerful Axiom of Replacement. This is deceptive, as Zermelo set theory with the Axiom of Rank is not essentially stronger than Zermelo set theory (it is possible to interpret the latter in the former), while the Axiom of Replacement makes Zermelo set theory far stronger.

Theorem (the Axiom of Foundation): Every set A has an element x such that x is disjoint from A.

Proof: Let r be the minimal rank in the inclusion order which includes an element of A as an element, and let $x \in r \cap A$. Each element of x belongs to a rank properly included in r, so x is disjoint from A.

The Axiom of Foundation is frequently (anachronistically) adjoined to Zermelo set theory as an additional axiom.

We observed above that the modern form of the Axiom of Infinity and the original form do not imply each other in the presence of the other axioms. They do imply each other in the presence of the Axiom of Rank. For the Axiom of Rank, combined with the existence of any infinite set, implies that there is a minimal infinite rank V_ω in the inclusion order, and both the Zermelo natural numbers and the von Neumann natural numbers are definable subsets of V_ω (since all of the elements of either are clearly of finite rank). It is also amusing to note that the Axioms of Pairing and Union can be omitted in the presence of the Axiom of Rank. $\{a, b\}$ can be derived using Separation as a subset of the power set of $r_a \cup r_b$ (this binary set union exists because it is actually one of the ranks r_a and r_b), and $\bigcup A$ can be derived using Separation as a subset of $\mathtt{TC}(A)$.

3.12.2 †Digression: Interpreting typed set theory as Mac Lane set theory

"Mac Lane set theory" is the version of Zermelo set theory with Separation replaced by Bounded Separation.

Mac Lane set theory can be interpreted in typed set theory with strong extensionality, using our entire universe of typed objects. We begin by postulating an operator J which is injective ($J(x) = J(y) \to x = y$) and sends type 0 objects to type 1 objects. An example of such an operator is the singleton operator ι. Any such J can be thought of as implemented by a function $\iota``V^{\mathbf{0}} \to V^{\mathbf{1}}$.

We now indicate how to extend the J operator to all types. If J is defined for type n objects we define $J(x^{\mathbf{n+1}})$ as $\{J(y^{\mathbf{n}}) \mid y^{\mathbf{n}} \in x^{\mathbf{n+1}}\}$. Briefly, $J(x) = J``x$. It is easy to see that J is injective on every type: we have $J(x) = J(y) \leftrightarrow x = y$, no matter what the common type of x and y. By the definition of J at successive types, we further have $J(x) \in J(y) \leftrightarrow x \in y$, no matter what the successive types of x and y.

In our interpretation of untyped set theory, we identify every object x of whatever type with each of its iterated images $J^n(x)$ under the J operator: in this way each type n is seen to be embedded in type $n+1$. If x is of type m and y is of type n, we have $x = y$ in the interpretation iff $J^n(x) = J^m(y)$ (note that both of these terms are of the same type $m + n$) and we have $x \in y$ in the interpretation iff $J^n(x) \in J^{m+1}(y)$ (in which the terms have successive types $m + n$ and $m + n + 1$). Notice that if x and y are of the same type n, $J^n(x) = J^n(y) \leftrightarrow x = y$, and if x and y are of successive types n and $n + 1$, $J^{n+1}(x) \in J^{n+1}(y) \leftrightarrow x \in y$: where equality and membership make sense in type theory, they coincide with equality and membership in the typed theory.

If x, y, z have types m, n, p, and we have $x = y$ and $y = z$, we have $J^n(x) = J^m(y)$ and $J^p(y) = J^n(z)$. Further applications of J to both sides of these formulas show that transitivity of equality works: $J^{n+p}(x) = J^{m+p}(y)$ and $J^{m+p}(y) = J^{m+n}(z)$ are implied by the previous equations and imply $J^{n+p}(x) = J^{m+n}(z)$, which in turn by injectivity of J implies $J^p(x) = J^m(z)$ which is the interpretation of $x = z$. Reflexivity and symmetry of equality present no difficulties. The substitution property of equality requires some technical detail for its verification which we do not (NOTE: yet) give here.

We verify that some of the axioms of Mac Lane set theory hold in this interpretation.

We discuss the Axiom of Extensionality. Suppose that x is of type n and y is of type $n+k$. If $k=0$ and x and y have the same elements, then $x=y$ by the axiom of extensionality of type theory. Otherwise, if for all z of type m we have $z \in x$ iff $z \in y$ in the interpreted theory, this means we have $J^n(z) \in J^m(x)$ iff $J^{n+k}(z) \in J^m(y)$, and further $J^{n+k}(z) \in J^{m+k}(x)$, whence $J^m(y) = J^{m+k}(x)$, whence $J^n(y) = J^{n+k}(x)$, whence $x = y$ in the interpretation, which is what is wanted.

Now we discuss the Axiom of Bounded Separation. We want to show the existence of $\{x \in A \mid \phi[x]\}$ in the untyped theory, where ϕ is a formula in membership and equality (it should not mention the predicate of typehood, which does not translate to anything in the language of type theory, though of course it may mention specific types) we suppose that every quantifier in $\phi[x]$ is restricted to a set. Assign referents to each free variable appearing in $\{x \in A \mid \phi[x]\}$, then assign each bound variable the type one lower than that assigned to the set to which it is restricted (A in the case of x, the bound on the quantifier in the case of quantified sets; if the bound is type 0, make the variable type 0 as well), then apply our interpretation of the untyped language in the typed language (adding applications of J to variables in such a way as to make everything well-typed). For example, $\{x \in A \mid x \notin x\}$ would become $\{x \in A \mid x \notin J(x)\}$, with x being assigned type one lower than that assigned to A (unless A was assigned a referent of type 0, in which case we would have $\{x \in J(A) \mid x \notin J(x)\}$). The resulting set abstract exists in our typed theory and has the right extension in the interpretation. If there were unbounded quantifiers in $\phi[x]$, there would be no way to interpret them in terms of our typed theory, which does not allow any way to quantify over objects of all types.

(NOTE: more axioms to be supplied. Rank will not necessarily hold here; the form of infinity which holds depends on the exact form of J. This development is more *ad hoc* and more closely related to the original form(s) of Zermelo set theory).

Something like this interpretation can also be carried out in the version of type theory with weak extensionality. We detail the modifications of the construction.

The operation J must be defined at atoms in each positive type. J is defined on type 0 as an injective operation raising type by 1, as above. If J is defined on type n objects, we define it on type $n+1$ *sets* as before: $J(x^{\mathbf{n+1}}) = J``(x^n)$. There are no more than $T|V^{\mathbf{n+1}}|$ elementwise images under J in type $n+2$: since $T|V^{\mathbf{n+1}}| < |\mathcal{P}(V^{\mathbf{n+1}})|$ by Cantor's theorem,

we can choose as many distinct further elements of $\mathcal{P}(V^{n+1})$, i.e., *sets* in V^{n+2}, as we need as images of the type $n+1$ atoms under J. The result $x \in y \leftrightarrow J(x) \in J(y)$ now holds only if y is a set, and for this reason we modify the interpretation of $x \in y$ in the untyped theory (where x and y have types m and n respectively in type theory) to $J^{n+1}(x) \in J^{m+1}(y)$; if x were of type 0 and y were an urelement of whatever type the original interpretation $J^n(x) \in J^m(y)$ would not work correctly.

3.12.3 †Translation between Type Theory and Set Theory

Importing results from type theory: We make a general claim here that mathematical results can be imported from type theory to untyped set theory. It is useful to give a uniform account of how such a general claim can be justified (which also makes it clear exactly what is claimed).

Just as we can translate the language of Zermelo set theory into the language of type theory in a way which makes the axioms true[10], so we can translate the language of type theory into the language of untyped set theory in a way which makes the axioms true – and so makes all the theorems true.

Let ϕ be a formula of the language of type theory mentioning n types. Let $X_0, X_1, \ldots X_{n-1}$ be a sequence of sets such that $\mathcal{P}(X_i) \subseteq X_{i+1}$ for each appropriate index i. The translation $(\phi)_X$ is defined as follows: each quantifier over type i is restricted to X_i. Each formula $x \in y$, where x is of type i and y is of type $i+1$ is translated as $x \in y \wedge y \in \mathcal{P}(X_i)$ (elements of $X_{i+1} - \mathcal{P}(X_i)$ are interpreted as urelements); formulas of the form $x = y$ are interpreted as $x = y$. Such a translation is also feasible if there is an infinite sequence with the same properties, but it is not a theorem of Zermelo set theory that there are such sequences. A specific version which we will write $[\phi]_X$ has $X_i = \mathcal{P}^{\mathbf{i}}(X)$ for a fixed set X: a nice feature of this version is that we can generate as many terms of the sequence as we need in a uniform way. It is straightforward to verify that as long as X_0 is infinite the translations of all axioms of type theory into the language of untyped set theory are true. It can further be noted that expressions T representing sets in the language of type theory will also have translations $(T)_X$ where X is a sequence or $[T]_X$ where X is a set.

This makes a wide class of mathematical assertions readily portable from type theory to set theory. For example, all of our assertions about cardinal and ordinal arithmetic have readily determined analogues in untyped set theory

We discuss how to transfer mathematical concepts and theorems from type theory to set theory.

[10]With qualifications discussed in section 3.7.1.

We have already seen that any formula of the language of type theory can be translated to a formula $[\phi]_X$ (where X is an infinite set) which asserts that ϕ holds in a model of type theory in which X is type 0, $\mathcal{P}(X)$ is type 1, and in general $\mathcal{P}^{\mathbf{n}}(X)$ is type n. $[\phi]_X$ is obtained by rereading membership and equality as the relations of the untyped theory and restricting each type n variable to $\mathcal{P}^{\mathbf{n}}(X)$. For each axiom ϕ of type theory (in each of its explicitly typed versions), it is straightforward to show that $[\phi]_X$ is a theorem of Zermelo set theory. So for any theorem ϕ of type theory we have $[\phi]_X$ a theorem of Zermelo set theory, and in fact we also have "for all infinite sets X, $[\phi]_X$" a theorem of Zermelo set theory.

Every object t we can define in the language of type theory has analogues t_X for each infinite set X. This presents an obvious problem (a stronger version of the ambiguity of type theory which our avoidance of type indices partially obscures). All our definitions of specific objects, with a few exceptions such as $\emptyset$, refer to different objects depending on the choice of the parameter X. For example the number $3^{\mathbf{n+2}}$ is implemented as $[\mathcal{P}^{\mathbf{n}}(X)]^3$, the set of all subsets of $\mathcal{P}^{\mathbf{n}}(X)$ with exactly three elements. Just which set this is varies with the choice of X (and n).

A possible conceptual problem with the theory of functions can be dispelled: in type theory, we can prove easily that the functions from a set A to a set B defined using Kuratowski pairs correspond precisely to those defined using Quine pairs (they are at different types but this ceases to be so inconvenient when we are translating to untyped set theory). So the question of which sets are the same size and which relations are isomorphic is settled in the same way no matter which pair definition one uses.

Nonetheless, the theory of cardinals and ordinals can be stated in untyped set theory as the theory of specific objects. Here we suppose that we use von Neumann ordinals as the implementation of ordinal numbers, and von Neumann initial ordinals as the implementation of cardinals. A sentence $(|A| = \kappa)_X$ asserts that A belongs to a certain cardinal κ_X. This translates to an assertion $|A| = \kappa$ in the language of untyped set theory, now not meaning $A \in \kappa_X$ but $A \sim \kappa$, where κ is the first von Neumann ordinal which is equinumerous with an element (and so with all elements) of κ_X. Further, it is important to note that for any cardinal $(\kappa^{\mathbf{n}})_X$ the von Neumann initial ordinal associated with it will be the same as the von Neumann initial ordinal associated with $(T(\kappa)^{\mathbf{n+1}})_X$: this gives a concrete meaning to our erstwhile intuitive feeling that κ and $T(\kappa)$ are in fact the same cardinal. Very similar considerations apply to order types $(\alpha)_X$ and corresponding von

Neumann ordinals α (and we get the analogous result that the ordinals $(\alpha^{\mathbf{n}})_X$ and $(T(\alpha)^{\mathbf{n+1}})_X$ correspond to the same von Neumann ordinal α). Further, nothing but technical points would differ if we used the Scott cardinals and ordinals here instead of the von Neumann cardinals and ordinals. Since we have a translation of ordinals and cardinals of type theory to ordinals and cardinals of the untyped set theory, we can translate the operations of addition and multiplication from type theory to untyped set theory directly. It might seem that we cannot translate cardinal exponentiation so directly, but here we observe that though $(\kappa^\lambda)_X$ is not always defined, it is always the case that $(T(\kappa)^{T(\lambda)})_X$ is defined (and will be $T(\kappa^\lambda)_X$ if the latter is defined); since the T operation is now understood to be the identity, we see that cardinal exponentiation is now a total operation. The way in which the definitions of cardinals and ordinals are transferred from type theory to set theory ensures that theorems of cardinal and ordinal arithmetic transfer as well. Notice that Cantor's Theorem now takes the form $\kappa < 2^\kappa$: there is no largest cardinal (from which it follows that there can be no universal set, as certainly $|V| \geq 2^{|V|}$ would hold; the argument from the untyped form of Cantor's theorem and the naive supposition that there is a universal set to a contradiction is called *Cantor's paradox*).

Although we have just defined operations of cardinal and ordinal arithmetic in terms of the interpreted type theory with X as type 0, it is perfectly possible to state definitions of these operations which do not depend on the notation $[\phi]_X$. The recursive definitions of operations of ordinal arithmetic are inherited directly by untyped set theory from type theory. The definitions of $|A| + |B|$ as $|(A \times \{0\} \cup B \times \{1\}|$, $|A| \cdot |B|$ as $|A \times B|$, and $|A|^{|B|}$ as $|A^B|$ work perfectly well in untyped set theory (always remembering that the set theoretical meaning of $|A|$, though not its mathematical function, is quite different). But the correspondence between the arithmetic of interpreted type theory and the arithmetic of untyped set theory is important in seeing that theorems can be relied upon to transfer fairly directly from type theory to set theory.

Results we have given above imply that certain statements which can be shown to be true in the version of Zermelo set theory interpreted in our type theory with strong extensionality are inconsistent with *ZFC*. We showed above that $\beth_\alpha$ does not exist for some α in these models (to be precise, if the cardinality of the set corresponding to type 0 is $\aleph_\beta$, we can prove that $\beth_{\beta+\omega}$ does not exist in that model (whereas in *ZFC* we have $|V_{\omega+\alpha}| = \beth_\alpha$ for each ordinal α, so all $\beth_\alpha$'s must exist)) However, there are models of Zermelo

set theory obtained from models of type theory with weak extensionality in which *ZFC* holds. This might seem not be possible since there is a sequence of sets $V^{\mathbf{i}}$ (the sets corresponding to the types) such that any cardinal is less than some $|V^{\mathbf{i}}|$ (since it is the cardinal of a set of some type): by Replacement it might seem that the countable sequence of $V^{\mathbf{i}}$'s would be a set (because it is the same size as the set of natural numbers), so its union would be a set, which would have cardinality larger than any $|V^{\mathbf{i}}|$. But this argument does not work, because there is no formula defining the sequence of $V^{\mathbf{i}}$'s (as there is in the models based on type theory with strong extensionality, where $V^{\mathbf{i+1}} = \mathcal{P}(V^{\mathbf{i}})$). Moreover, we will apply simple model theory below to show that for any model of *ZFC* there is a model obtainable from a model of type theory with weak extensionality in which the same statements of the language of set theory are true [that is a very convoluted sentence, I know].

The serious difference in power between untyped set theory and typed set theory has to do with the ability to quantify over the entire universe. This is just a difference in what we can *say* if we use Bounded Separation, but if we adopt the full axiom of Separation we can define sets in terms of global facts about the universe. This is best indicated using an example.

Theorem: For each natural number n, there is a unique sequence s of sets with domain the set of natural numbers $\leq n$ such that $s_0 = \mathbb{N}$ and for each $i < n$, $s_{i+1} = \mathcal{P}^n(\mathbb{N})$.

Proof: Prove this by mathematical induction. The set of natural numbers n for which there is such a sequence s clearly includes 0 ($s = \langle 0, \mathbb{N}\rangle$) and if it includes k will also include $k+1$ (if s works for k, $s \cup \langle k+1, \mathcal{P}(s_k)\rangle$ works for $k+1$).

Discussion: In type theory with base type countable, sets interpreting these sequences do not all exist in any one type, so no assertion of type theory can even express the fact that they all exist. This statement is of course very badly typed, but a similar assertion would be the statement that there is a sequence of cardinals such that $s_0 = \aleph_0$ and $s_{i+1} = 2^{s_i}$ for each i, and this would present the same problem: in type theory with base type countable, the sequence $\beth_0, \beth_1 \beth_2, \ldots$ is not entirely present in any one type. The mere statement of the theorem cannot be expressed in type theory because the quantifier over sequences s is not bounded in a set (and for this same reason this theorem cannot be proved using

Bounded Separation: the subset of the natural numbers which needs to be shown to be inductive cannot be shown to exist).

Chapter 4

Logic

4.1 Formalization of Syntax and Substitution

In this section we discuss the representation of bits of syntax (formulas and terms) by mathematical objects. We will thereafter identify the syntactical objects of our language with these mathematical objects.

An obvious way to do this would be to represent ASCII characters by natural numbers, then represent character strings as functions from finite initial segments of $\mathbb{N}$ to ASCII characters. But the definition of formal operations on syntax with this definition would be inconvenient.

Our representation will be typically ambiguous, as with all our representations of mathematical objects in type theory: syntactical objects will exist in all types above a certain minimum type (which we really will not care about determining). Though we work in type theory it should be clear how the same construction could be done in Zermelo set theory.

4.2 A toy example (calculator arithmetic)

Our full type theory has a quite complex language, so we provide a preliminary example of construction of a formal language within our type theory and definition of semantics for it (intended values for all the expressions of the language as represented within our theory). The objects we use to represent expressions of calculator arithmetic will all be of the same type, some fixed $n + 2$.

The language we consider is the language of calculator arithmetic.

Each individual digit is assigned its usual meaning (0,1,2,3,4,5,6,7,8,9).

Strings of digits are to be assigned their usual meanings: if D has been assigned a value x already, and d is a digit whose value r we already of course know, the value of Dd (understood as a string concatenation) will be $10x+r$.

General expressions will include all the strings of digits and all sums and products of expressions. So we expect $(102 + 5) \cdot 13$ to be an expression, for example.

Trying to represent our symbols as strings is certainly possible, but would require reasoning about mathematical representations of parentheses which would be quite unpleasant. We take a different tack, which handles grouping without parentheses.

digits: Each digit $n \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ is represented by $(0, n)$. Using quotes, we define 'n' as $(0, n)$.

base ten numbers: A base ten number with one digit n is represented by the digit $(0, n)$. A base ten number N whose last digit is n and which has more than one digit is represented by $(1, D, (0, n))$, where D is the representation of $\frac{N-n}{10}$. So for example 5 is coded by $(0, 5)$, 12 is coded by $(1, (0, 1), (0, 2))$ and 365 is coded by $(1, (1, (0, 3), (0, 6)), (0, 5))$. Using quotes, if '$d$' is a digit (so d is a known small number) and 'D' is a decimal numeral, 'Dd' (the string obtained by appending 'd' to 'D') is defined as $(1, `D', `d')$.

arithmetic expressions: A base ten number by itself is an arithmetic expression.

If E and F are arithmetic expressions, $(2, E, F)$ represents the formal sum of these two notations and $(3, E, F)$ represents the product of these two notations.

Using quotes, if 'E' is a calculator expression and 'F' is a calculator expression, we define '$(E{+}F)$' as $(2, `E', `F')$ and '$(E{\cdot}F)$' as $(3, `E', `F')$.

The translation of the calculator notation "$(102 + 5) \cdot 13$" will then be $(3, (2, (1, (1, (0, 1), (0, 0)), (0, 2)), (0, 5)), (1, (0, 1), (0, 3)))$.

The point here is that the individual notations are objects internal to our type theory, rather than symbols. An alternative way to do this would be to postulate something like character strings as objects of our theory, but it is instructive that we do not need any new primitive ideas to implement this.

It is important to note that these notations represent pieces of notation, not numbers. "2+3" (a piece of notation) is not the same thing as "5" or "3+2", though these pieces of notation represent the same numbers. And all three of these are different complex pairs.

It is not enough for us to be able to represent each individual piece of calculator notation. We want to be able to say that something is a piece of calculator notation internally to our mathematical language. We define the sets corresponding to the categories of notation we have discussed.

the set of digits: The set D of digits is easy to define: this is the set $\{0\} \times \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

the set of base ten numbers: We will define the set T of base ten numbers.

We say that a set I is T-inductive iff $D \subseteq I$ and $\{1\} \times I \times D \subseteq I$. We call the set of all T-inductive sets $\mathcal{T}$ and define T as $\bigcap \mathcal{T}$, the collection of all objects which belong to every T-inductive set.

Certainly the collection of all base ten numbers is T-inductive. And any T-inductive set (a collection I which contains all the digits and has the property that any triple $(1, x, d)$ where x is in the collection and d is a digit) must contain all the base ten numbers. An example of a T-inductive set which is not the collection of all base ten numbers (other than the trivial example, the universe), is the collection $J = (\{0\} \times V) \cup (\{1\} \times V \times D)$ of all things which are either a pair with first component 0 or a triple with first component 1 and last component a digit.

the set of calculator expressions: We define the set E of calculator expressions.

We say that a set I is E-inductive iff $T \subseteq I$ and $\{2\} \times I \times I \subseteq I$ and $\{3\} \times I \times I \subseteq I$. We call the set of all E-inductive sets $\mathcal{E}$, and we define E as $\bigcap \mathcal{E}$, the collection of all objects which belong to any E-inductive set. It should be clear both that the collection of calculator expressions should be E-inductive, and that any E-inductive set that we define will contain all the calculator expressions.

Finally, of course, the most interesting thing about a piece of notation is *what it means*. We will define a function $v : E \to \mathbb{N}$ which will represent the natural number value which we expect a piece of calculator notation to denote (what display do you get when you type this notation into the calculator?). A function like v is called a "valuation".

We list conditions which we expect to hold of v.

valuation of digits: $v((0, n)) = n$ about sums up our expectations. This is the simple case. So we expect $((0, n), n) \in v$ for each $n \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

valuation of base ten numerals: A base ten numeral which is a digit we already know how to handle. A base ten numeral of the form $(1, D, (0, n))$ will have $v((1, D, (0, n)) = (10 \cdot v(D)) + n$. So if $(D, x) \in v$, we expect $((1, D, (0, n)), 10 \cdot x + n) \in v$.

valuation of calculator expressions: A calculator expression which is a base ten numeral I already know how to handle. We expect $v((2, e, f)) = v(e) + v(f)$ and $v((3, e, f)) = v(e) \cdot v(f)$. We express this a little differently: if $(e, x) \in v$ and $(f, y) \in v$, we expect $((2, e, f), x + y) \in v$ and $((3, e, f), x \cdot y) \in v$.

the appropriate kind of inductive set: A set I is v-inductive iff $((0, n), n) \in I$ for each $n \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, and if $(D, x) \in I$ and $D \in T$ and $x \in \mathbb{N}$ we have $((1, D, (0, n)), 10 \cdot x + n) \in I$, and if $(e, x) \in I$ and $(f, y) \in I$, and $x, y \in \mathbb{N}$, we have $((2, e, f), x + y) \in I$ and $((3, e, f), x \cdot y) \in I$.

the definition of v: Define $\mathcal{V}$ as the collection of all v-inductive sets and define v as $\bigcap \mathcal{V}$, the collection of objects which belong to every v-inductive set. This way of constructing a function may make us queasy, and should remind us of the proof of the Iteration Theorem. An informal argument that this is correct should exploit the observation in effect

already made that our first three points show that v, considered as a set of pairs, actually is v-inductive – and further that any v-inductive set should actually contain all the pairs in v, so their intersection is exactly v.

Exercises

The first two problems may be all that you do, but I do invite you to think about the two harder problems which follow.

1. For each of the following nested pair expressions, determine whether it actually is a mathematical representation of a calculator expression (including digits and base ten numbers), and if it is one report its value. Show steps in your calculation.
   (a) $(2,(1,(0,1),(0,3)),(0,5)$
   (b) $(1,(0,1),(1,(0,3),(0,4)))$
   (c) $(3,(2,(0,5),(0,3)),(1,(0,1),0,0))$
   (d) $(1,(2,(0,2),(0,3)),(0,5))$
2. Write the nested pair expressions which represent the following expressions of calculator arithmetic. You do not have to compute values.
   (a) $5+4$
   (b) $(2+3)+4$
   (c) $2+(3+4)$
   (d) $15 \cdot 234$
3. This is a challenge problem: determine why I need to say "$d \in T$" in the clause "if $(d,x) \in I$ and $d \in T$ and $x \in \mathbb{N}$ we expect $((1,d,(0,n)),10 \cdot x+n) \in I$" (hint: give an example of an illegal expression (something not in E) at which we would be forced to evaluate v if we did not include this condition). One of the parts of the first problem is relevant!
4. Another challenge problem: show that the set of pairs $V \times \mathbb{N}$ is v-inductive – this shows that every value of v is a natural number. ($V \times \mathbb{N}$ is the set of all ordered pairs whose second component is a natural number).

4.3 A formal syntax for our type theory

We initially give a recursive definition of notation taken from logic and set theory as mathematical objects.

We begin with variables. The triple $\langle 0, m, n\rangle$ will represent a bound variable $x_n^{\mathbf{m}}$ and the triple $\langle 1, m, n\rangle$ will represent a free variable (or "constant") $a_n^{\mathbf{m}}$ for natural numbers m, n. The reasons why we want bound and free variables will become evident later. That is, we define 'x_n^m' as $\langle 0, m, n\rangle$ and 'a_n^m' as $\langle 1, m, n\rangle$.

The triple $\langle 2, n, t\rangle$, where t is a term, will represent the sentence $P_n(t)$ (P_n being a unary predicate). The quadruple $\langle 3, n, t, u\rangle$ will represent the sentence $t\, R_n\, u$ (R_n being a binary predicate (logical relation)). We read $\langle 3, 0, t, u\rangle$ as $t \subseteq u$ and $\langle 3, 1, t, u\rangle$ as $t = u$. That is, we define '$P_n(t)$' as $\langle 2, n, \text{'}t\text{'}\rangle$ and '$t\, R_n\, u$' as $\langle 3, n, \text{'}t\text{'}, \text{'}u\text{'}\rangle$.

The triple $\langle 4, n, t\rangle$ (t being a term) stands for $F_n(t)$ (F_n being a function symbol). The quadruple $\langle 5, n, t, u\rangle$ (t and u being terms) stands for $t\, O_n\, u$, O_n being a binary function (operation) symbol. That is, '$F_n(t)$' is defined as $\langle 4, n, \text{'}t\text{'}\rangle$ and '$t\, O_n\, u$' is defined as $\langle 5, n, \text{'}t\text{'}, \text{'}u\text{'}\rangle$.

We reserve F_0 and F_1 to stand for the projection operators, and O_0 to stand for the ordered pair.

Note that all predicate and function symbols are typically ambiguous (can be used with arguments of many types). Binary relation symbols are assumed to be type level and functions are assumed to have one or both inputs and their output all of the same type.

The triple $\langle 6, n, t\rangle$ represents $\iota^{\mathbf{n}}(t)$ and the triple $\langle 7, n, t\rangle$ represents $\bigcup^{\mathbf{n}} t$. That is, '$\iota^{\mathbf{n}}(t)$' is defined as $\langle 6, n, \text{'}t\text{'}\rangle$ and '$\bigcup^{\mathbf{n}} t$' is defined as $\langle 7, n, \text{'}t\text{'}\rangle$. [It is important to note that we intend in our semantics to extend the union operation so that $\bigcup\{x\}$ is equal to x even if x is an atom.]

Note that we can now represent $t \in u$ as $\{t\} \subseteq u$. The reason why we choose this apparently odd representation of the membership relation is that we can then allow all formal relations in the syntax to be type-level, which makes the definition of the type of an expression simpler.

The quadruple $\langle 8, n, v, \phi\rangle$ (where ϕ is a formula and v is a bound variable) is read $(Q_n v.\phi)$, where Q_n is a quantifier. We reserve Q_0 as $\exists$ and Q_1 as $\forall$. That is, '$(Q_n v.\phi)$' is defined as $\langle 8, n, \text{'}v\text{'}, \text{'}\phi\text{'}\rangle$.

We briefly recall that the symbol $(\epsilon x.\phi)$ represents an arbitrarily chosen element of $\{x \mid \phi\}$ for each ϕ, and a default object (which we could take to be $\emptyset$ if the type of the expression is positive) if $\{x \mid \phi\}$ is empty: the Axiom

of Choice allows us to assume that in each type we have a suitable choice function which picks an element from each nonempty set.

The quadruple $\langle 9, n, v, \phi\rangle$ (where ϕ is a formula and v is a bound variable) represents a term $(B_n v.\phi)$ constructed by binding on a formula. We read $\langle 9, 0, v, \phi\rangle$ as $(\epsilon v.\phi)$, the Hilbert symbol. That is, '$(\epsilon v.\phi)$' (in particular) is defined as $\langle 9, 0, `v', `\phi'\rangle$. Note that '$\{v \mid \phi\}$' can be read as '$(\epsilon A.(\forall v.\{v\} \subseteq A \leftrightarrow \phi))$', where A is the first variable of appropriate type not found elsewhere in the expression.

Alternatively, we could allow $\langle 9, 0, `v', `\phi'\rangle$ to represent $\{x \mid \phi\}$, in which case our rules for typing this expression would be different; but the Hilbert symbol is actually very useful in formal logic, though less familiar to us, and we prefer to provide it as a syntactical primitive.

The pair $\langle 10, \phi\rangle$ represents $\neg\phi$. The triple $\langle 11, \phi, \psi\rangle$ represents $\phi \vee \psi$. That is, '$\neg\phi$' is defined as $\langle 10, `\phi'\rangle$ and '$\phi \vee \psi$' is defined as $\langle 11, `\phi', `\psi'\rangle$. We could equally well use the construction $\langle 11, n, \phi, \psi\rangle$ and provide ourselves with a potentially infinite supply of binary propositional connectives: $\langle 11, n, `\phi', `\psi'\rangle$ would be taken to code '$\phi \oplus_n \psi$', and we would reserve $\oplus_0, \oplus_1, \oplus_2, \oplus_3$ for $\wedge, \vee, \rightarrow, \leftrightarrow$.

The above is not precisely mathematical as it relates mathematical objects to pieces of notation. We proceed to develop a thoroughly mathematical account of syntax and semantics using this informal account as motivation. For readability, we will allow ourselves to use quoted terms and formulas much of the time.

Definition: This is a nonce definition. A syntactical pair of sets is a pair of sets $\langle T, F\rangle$ with the following properties, motivated by the idea that T is an approximation to the set of terms and F is an approximation to the set of formulas.

1. For any natural numbers m, n, $\langle 0, m, n\rangle$ and $\langle 1, m, n\rangle$ belong to T. Objects $\langle 0, m, n\rangle$ are called bound variables.
2. For any natural number n and any $t \in T$, $\langle 2, n, t\rangle \in F$.
3. For any natural number n and $t, u \in T$, $\langle 3, n, t, u\rangle \in F$.
4. For any natural number n and $t \in T$, $\langle 4, n, t\rangle$, $\langle 6, n, t\rangle$, $\langle 7, n, t\rangle \in T$
5. For any natural number n and $t, u \in T$, $\langle 5, n, t, u\rangle \in T$.
6. For any natural number n, , bound variable v, $\phi \in F$, and $t, u \in T$, $\langle 8, n, v, \phi\rangle \in F$, and $\langle 9, n, v, \phi\rangle \in T$.

7. For any $\phi, \psi \in F$, $\langle 10, \phi\rangle \in F$ and $\langle 11, \phi, \psi\rangle \in F$.

Definition: A *formal term set* is any set which is the first projection T of a syntactical pair $\langle T, F\rangle$. A *formal proposition set* is any set which is the second projection F of a syntactical pair $\langle T, F\rangle$. A *formal term* is an object which belongs to all formal term sets. A *formal proposition* is an object which belongs to all formal proposition sets.

Theorem: If $\mathcal{T}$ is the set of all formal terms and $\mathcal{F}$ is the set of all formal propositions, then $\langle \mathcal{T}, \mathcal{F}\rangle$ is a syntactical pair of sets.

The two sets $\mathcal{T}$ and $\mathcal{F}$ are defined by mutual recursion. It is natural to prove theorems about formal terms and propositions using structural induction. We will write formal terms and propositions using ordinary typography, and in fact to the best of our ability forget the intricacies of numerals and pairing that underly the formal definition (particularly since the details are largely arbitrary and could be changed wholesale without affecting the subsequent development).

Terms have type, and considerations of type determine that some terms are ill-formed. $x_n^{\mathbf{m}}$ and $a_n^{\mathbf{m}}$ have type m. $F(t)$ has the same type as t. $t\,O\,u$ has type m iff t and u have the same type m and is ill-typed otherwise. $(\epsilon x_n^{\mathbf{m}}.\phi)$ (this is the Hilbert symbol) has type m. A formula $t\,R\,u$ will only be considered well-formed if t and u have the same type. If t has type n, $\iota^{\mathbf{k}}(t)$ has type $n+k$ and $\bigcup^{\mathbf{k}}(t)$ has type $n-k$ if $n \geq k$ and is considered ill-formed otherwise. These clauses are enough to determine the typing (and well-formedness) of all terms and formulas by recursion.

Now we give the formal definition of substitution. We define $u[t/x_i]$ (the result of replacing x_i with t in the term u) and $\phi[t/x_i]$ (the result of replacing x_i with t in the formula ϕ) at the same time. Here we leave off the type index: the type requirement is that t and x_i have the same type.

1. $x_j[t/x_i]$ is defined as t if $i = j$ and as x_j otherwise.
2. $a_j[t/x_i]$ is defined as a_j.
3. $F(u)[t/x_i]$ is defined as $F(u[t/x_i])$.
4. $(u\,O\,v)[t/x_i]$ is defined as $u[t/x_i]\,O\,v[t/x_i]$.
5. $(Bx_j.\phi)[t/x_i]$ is defined as $(Bx_k.\phi[x_k/x_j][t/x_i])$, where x_k is the first variable not found in $(Bx_j.\phi)[t/x_i]$. The only B that we use is the Hilbert symbol (which we use to express set abstraction as indicated above).
6. $P(u)[t/x_i]$ is defined as $P(u[t/x_i])$.
7. $(u\,R\,v)[t/x_i]$ is defined as $u[t/x_i]\,R\,v[t/x_i]$.
8. $(Qx_j.\phi)[t/x_i]$ is defined as $(Qx_k.\phi[x_k/x_j][t/x_i])$, where x_k is the first variable not occurring in $(Qx_j.\phi)[t/x_i]$.
9. $(\neg\phi)[t/x_i]$ is defined as $\neg\phi[t/x_i]$ and $(\phi \vee \psi)[t/x_i]$ is defined as $\phi[t/x_i] \vee \psi[t/x_i]$.

To justify that this definition works takes a little thought. The notion of length of a term or formula can be defined by a natural recursion (we do not give the mind-numbing details here). Then observe that the substitution of t for x_i in any given formula P is may be defined in terms of other substitutions supposed already defined, but these are always substitutions into strictly shorter formulas.

Our formulation of syntax differs from usual formulations in defining a single universal formal language, which is specifically adapted to the needs of type theory, though it can also be used for single-sorted first order theories. The adaptation to first-order theories is straightforward: simply do not use variables of type other than zero or the singleton or union operations. The language would need to be extended for more complicated multi-sorted theories (more complicated type theories): we will not discuss this. The language

could be extended with n-ary predicate and function symbols for $n > 2$, of course. It can obviously be cut down by specifying limited collections of constants, unary and binary predicate symbols, and unary and binary function symbols.

4.3.1 Exercises

1. Using the definitions of formal syntax above, write out the mathematical object coding the formula

$$(\forall x_1^3.(\exists x_2^3.x_1^3\, R_2\, x_2^3)).$$

2. What is the term or formula coded by

$$\langle 8, 1, \langle 0, 0, 1\rangle, \langle 3, 2, \langle 0, 0, 1\rangle, \langle 1, 0, 1\rangle\rangle\rangle?$$

3. Formalize the process of substituting $2 + 3$ for x in the sentence $(x + y) + z = x + (y + z)$, starting with the formal expression

$$((x + y) + z = x + (y + z))[(2 + 3)/x]$$

and proceeding agonizing step by agonizing step as I did in class. In words, this means "replace x with $2+3$ in the expression $(x+y)+z = x + (y + z)$" and should give just the result you expect. But anyone doing formal syntax should expand something like that once (grin).

4. The expression $(\forall x : (\forall y : c(x + y) = cx + cy))$ is true of course. Suppose we replace c with y. Then we might think we are saying $(\forall x : (\forall y : y(x+y) = yx+yy))$, that is, $(\forall x : (\forall y : y(x+y) = yx+y^2))$. Now certainly this is true, but it is not really the statement which I mean when I say to replace c with y in the original statement. Can you see why not? What would the offical result of this substitution look like (you do not need to write formal expansions, just write out the intended sentence, using whatever choice of variables seems reasonable).

4.4 Formalization of Reference and Satisfaction

In this section we define the notions of *meaning* and *truth*. That is, given an interpretation of the nonlogical symbols of our language, we show how to formally define the referent of each term and the truth value of each formula, mod assignments of values to all variables.

We first need to set the stage. A domain of objects is needed to support our interpretation. In fact, we supply a sequence D_n of domains, one for each $n \in \mathbb{N}$, with D_n intended to be the collection of type n objects.

Note that all the sets D_n are actually of the same type in the sense of our working type theory. If we restrict our language to the first-order as indicated above, we only need a single domain D. We will use M to represent the type of the elements of D_n (the type of the objects that terms of our language stand for). We will stipulate that the terms of our language are also of type M. It follows that each of the domains D_n is of type $M+1$ and the sequence D is of type $M+2$ (a map sending type $M+1$ natural numbers to type $M+1$ sets).

We associate a value $a_i^n \in D_n$ with each constant $a_i^{\mathbf{n}}$. With each unary predicate P_i we associate a set $P_i^n \subseteq D_n$ for each n (because our language is typically ambiguous we need an interpretation of each predicate over each type). With each binary relation symbol R_i we associate a set $R_i^n \subseteq D_n \times D_n$ for each n. Similarly each unary function symbol F_i is associated with functions $F_i^n : D_n \to D_n$, and each binary operation symbol O_i with functions $O_n : D_n^2 \to D_n$. An injective map $\iota_{n+1} : D_n \to D_{n+1}$ is provided for each n, and a map $\bigcup_n : D_{n+1} \to D_n$ with the property $\bigcup_n(\iota_{n+1}(x)) = x$ for each $x \in D_n$. We define $\iota_{n,k}$ as the identity map on D_n if $n = k$ and for $n < k$ define $\iota_{n,k}$ as $\iota_k \circ \iota_{n,k-1}$: the operation $\iota_{n,k}$ implements the representation of $(k-n)$-fold singletons of type n objects. We define $\bigcup_{n,k}$ as the identity map on D_n if $n = k$ and for $n > k$ define $\iota_{n,k}$ as $\bigcup_k \circ \bigcup_{n,k+1}$: the operation $\bigcup_{n,k}$ implements the representation of $(n-k)$-fold unions of type n objects. It is useful to note that I have indexed these operations so that the index (the second one, if there are two) is the object language type of the output. (The existence of these latter maps imposes requirements on the sequence of sets D_n: the sets in the sequence must be of increasing size). To support the Hilbert symbol we provide a function H_n from nonempty subsets of D_n for each n: $H_n(A) \in 1$ for all A; if $A \subseteq D_n$, then $H_n(A) \subseteq A$ if $A \neq \emptyset$; $H_n(\emptyset)$ is

defined and belongs to ι"D_n but is otherwise unspecified.

A *structure* for our formal language is determined by a map D sending a possibly proper initial segment of the natural numbers to domains D_n, "singleton" and "union" maps $\iota^{\mathbf{n+1}} : D_n \to D_{n+1}$ and $\bigcup^{\mathbf{n}} : D_{n+1} \to D_n$ as above, modified choice functions H_n as above (if the Hilbert symbol is to be used), and some partial functions implementing constants, predicates and functions as indicated above: where m, n are natural numbers, $A(m, n)$ will be the element a_n^m of D_n used as the referent of $a_n^{\mathbf{m}}$, $P(m, n)$ will be the subset P_n^m of D_m intended to be the extension of the predicate P_n in type m, $R(m, n)$ will be the subset R_n^m of $D_m \times D_m$ intended to be the extension of the logical relation R_n, $F(m, n)$ is the element F_n^m of $D_m^{D_m}$ (recall that B^A is the set of functions from A to B) representing the action of the function symbol F_n in type m, and $O(m, n)$ is the element O_n^m of $D_m^{D_m \times D_m}$ representing the action of the operation symbol O_n in type m. The length of the domain sequence and the domain of the partial function determine the subset of our universal language which is used in the obvious way.

The binding constructions used in the discussion which follows are limited. The only term construction binding propositions we provide is the Hilbert symbol $(\epsilon x.\phi[x])$ which may be read "an x such that $\phi[x]$ if there is one (chosen in an unspecified manner if there are more than one) or a default object if there is no such x". All definable term binding constructions (including the set builder notation) can be defined in terms of the Hilbert operator. The only quantifiers we provide are the usual ones (which can in fact also be defined in terms of the Hilbert operator!). It is not difficult to extend the discussion to general binders, but it would further complicate already very elaborate recursive definitions.

A possibly partial function E on variables such that $E(x_i^{\mathbf{n}}) \in D_n$ for each variable $x_i^{\mathbf{n}}$ in the domain of E is called an *environment*. If E is an environment we define $E[d/x_i^{\mathbf{n}}]$ as the environment which sends $x_i^{\mathbf{n}}$ to d and agrees with E everywhere else (this may be an extension of E if E is not defined at $x_i^{\mathbf{n}}$). Notice that each environment is an object of type $M + 1$. [If we restricted ourselves to finite partial functions as environments, it is possible to use type M objects built with ordered pairing.]

We will now recursively define functions $\mathcal{R}$ and $\mathcal{V}$ (named with "reference" and "valuation" in mind). These functions take two arguments, an environment and a term: strictly speaking, because of typing, they need to be type $M + 2$ functions taking an environment and the singleton of a term as arguments. They are partial functions: they are sometimes undefined. Strictly

speaking, these functions are defined relative to a structure and would be written $\mathcal{R}_{\mathcal{S}}$ and $\mathcal{V}_{\mathcal{S}}$ if we wanted to explicitly specify a structure S we were working with. We use the informal notation a_i^n for $A(n,i)$, P_i^n for $P(n,i)$, and so forth. The domains of these functions are restricted to the language appropriate to the structure (and further restricted depending on the extent to which E is partial).

We define $\chi(\phi)$ as 1 if ϕ is true and 0 if ϕ is false. Note that χ is a truly weird operation taking a sentence of our metalanguage to a number; all uses of this device can actually be eliminated, but it is a convenience. It *is* possible to define $\chi(\phi)$ quite honestly, as $\{x \mid (\phi \wedge x \in 1) \vee (\neg\phi \wedge x \in 0)\}$.

Now for the horrible recursive definition.

1. $\mathcal{R}(E, x_i^{\mathbf{n}}) = E(x_i^{\mathbf{n}})$ (if this is defined).
2. $\mathcal{R}(E, a_i^{\mathbf{n}}) = A(n, i)$.
3. $\mathcal{R}(E, F_i(t))$ is defined as $F(n, i)(\mathcal{R}(E, t))$, where n is the type of t (as long as $\mathcal{R}(E, t)$ is defined).
4. $\mathcal{R}(E, \iota^{\mathbf{k}}(t))$ is defined as $\iota_{n,n+k}(\mathcal{R}(E, t))$, where n is the type of t, as long as the embedded reference is defined.
5. $\mathcal{R}(E, \bigcup^{\mathbf{k}}(t))$ is defined as $\bigcup_{n,n-k}(\mathcal{R}(E, t))$, where n is the type of t, as long as the embedded reference is defined.
6. $\mathcal{R}(u\, O_i\, v)$ is defined as $O(n, i)(\mathcal{R}(E, u), \mathcal{R}(E, v))$ just in case $\mathcal{R}(E, u)$ and $\mathcal{R}(E, v)$ are defined and u and v have the same type n.
7. $\mathcal{V}(E, P(u))$ is defined as $\chi(\mathcal{R}(E, u) \in P(n, i))$, where n is the type of u, as long as the embedded reference is defined.
8. $\mathcal{V}(E, (u\, R_i\, v)$ is defined as $\chi(\langle \mathcal{R}(E, u), \mathcal{R}(E, v)\rangle \in R(n, i))$, as long as the embedded references are defined and u and v have the same type n.
9. $\mathcal{V}(E, \neg\phi)$ is defined as $\chi(\neg(\mathcal{V}(E, \phi) = 1))$, and $\mathcal{V}(E, \phi \vee \psi)$ is defined as $\chi(\mathcal{V}(E, \phi) = 1 \vee \mathcal{V}(E, \psi) = 1)$, as long as the embedded valuations are defined.
10. $\mathcal{V}(E, (Qx_j^{\mathbf{n}}.\phi))$ is defined as $\chi((Qd \in D_n.\mathcal{V}(E[d/x_i^{\mathbf{n}}], \phi) = 1))$, where Q is either $\exists$ or $\forall$, as long as the embedded valuation is defined. Please notice that the quantifiers on the left side of the definition are in quotes and on the outside are real quantifiers of our metalanguage (restricted to the appropriate D_n).
11. $\mathcal{R}(E, (\epsilon x_i^{\mathbf{n}}.\phi))$ is defined as the sole element of $H_n(\{d \in D_n \mid \mathcal{V}(E[d/x_i^{\mathbf{n}}], \phi) = 1\})$ if the valuation is defined.

Notice as with substitution that the reference and valuation functions are defined recursively. Reference and valuation for a particular term or formula may appeal to reference or valuation for another formula or term, but always a strictly shorter one.

Although our language is restricted for convenience in framing these definitions, the full language of type theory is supported with suitable definitions. If equality and subset relations are primitive, we define $x \in y$ as $\iota(x) \subseteq y$, $\phi \rightarrow \psi$ as $\neg\phi \vee \psi$, $\phi \wedge \psi$ as $\neg(\neg\phi \vee \neg\psi)$, $\phi \leftrightarrow \psi$ as $\phi \rightarrow \psi \wedge \psi \rightarrow \phi$, and $\{x \mid \phi\}$ as $(\epsilon A.(\forall x.x \in A \leftrightarrow \phi))$.

A further technical note is that $\mathcal{R}_S$ and $\mathcal{V}_S$ are lateral operations: they actually take a type $M+1$ environment and a type M term to a type M term or truth value. They can of course be transformed into sets by encasing the second argument and the value produced in singleton brackets, but we will not do this. We *will* suppose that we have done this in any context where we presuppose that we have identified an $\mathcal{R}_S$ and $\mathcal{V}_S$ as actual objects of our type theory.

4.4.1 Exercises

1. Use the definitions of reference and satisfaction to evaluate the following expressions, if $D_0 = \{1, 2, 3\}$ and the following information about the environment and interpretation is given. Notice that we really do not need to worry about types in this example.

   $A(0, 1) = 3$ (that is, the intended referent of a_1^0 is 3).

   $P(0, 1) = \{1, 2\}$

   $R(0, 1) = \{\langle 1, 1\rangle, \langle 2, 2\rangle, \langle 3, 3\rangle\}$ (the equality relation).

   $E(x_n^0) = 1$ for all n (the environment E assigns every type 0 variable the value 1).

   Show the reasoning behind your evaluation in detail. The intended evaluations are quite obvious: the point is to show that the nasty definitions in the notes actually get us there, so detail must be seen. This is an exercise in step by step unpacking of definitions.

   (a) $\mathcal{R}(E, a_1^0)$
   (b) $\mathcal{R}(E, x_5^0)$
   (c) $\mathcal{V}(E, P_1(a_1^0))$
   (d) $\mathcal{V}(E, P_1(x_1^0))$
   (e) $\mathcal{V}(E, x_2^0\, R_1\, a_1^0)$
   (f) $\mathcal{V}(E, x_2^0\, R_1\, x_5^0)$
   (g) $\mathcal{V}(E, (\exists x_2^0 . x_2^0\, P_1\, a_1^0))$

2. Why didn't we define $\mathcal{V}(E, (\forall x.\phi)$ as "χ(for all terms t, $\mathcal{V}(E, \phi[t/x])$=1)"? Such a scheme, called "substitutional quantification", does have fans. But it is not equivalent to our scheme. Can you see why? Hint: it is making a very strong assumption about the capabilities of our formal language.

4.5 Formal Propositional Sequent Calculus

We introduce sequent notation.

Definition: A *sequent* is an ordered pair $\langle \Gamma, \Delta \rangle$ of finite sets of formulas. We write sequents $\Gamma \vdash \Delta$. The set $\{A\}$ (where A is a formula) is simply written A in a sequent; the set $\Gamma \cup \{A\}$ is written Γ, A; notation for the empty set is omitted.

Definition: A sequent $\Gamma \vdash \Delta$ is *valid* iff every interpretation under which $\mathcal{V}$ is defined for all elements of Γ and Δ [we will presume this condition for all interpretations and sequents hereinafter] and under which $\mathcal{V}$"$\Gamma \subseteq \{1\}$ has $1 \in \mathcal{V}$"Δ (every interpretation which makes *all* statements in Γ true makes *some* statement in Δ true).

Lemma: $\Gamma, A \vdash \Delta, A$ is a valid sequent for any formula A and sets Γ and Δ.

Lemma: $\Gamma, \neg A \vdash \Delta$ is a valid sequent iff $\Gamma \vdash A, \Delta$ is a valid sequent.

Lemma: $\Gamma \vdash \neg A, \Delta$ is a valid sequent iff $\Gamma, A \vdash \Delta$ is a valid sequent.

Lemma: $\Gamma, A \vee B \vdash \Delta$ is a valid sequent iff both $\Gamma, A \vdash \Delta$ and $\Gamma, B \vdash \Delta$ are valid sequents. Note that this is a formalized version of the strategy of proof by cases.

Lemma: $\Gamma \vdash A \vee B, \Delta$ is a valid sequent iff $\Gamma \vdash A, B, \Delta$ is a valid sequent.

We introduce a weaker notion of valuation appropriate when we are considering propositional logic only.

Definition: A *propositional valuation* is a partial function $\mathcal{V}$ which sends each formula in its domain to either 0 or 1, and which sends any formula $\neg\phi$ to $1 - \mathcal{V}(\phi)$ and any formula $\phi \vee \psi$ to $\mathcal{V}(\phi) + \mathcal{V}(\psi) - \mathcal{V}(\phi) \cdot \mathcal{V}(\psi)$ (in each case iff the valuations of subformulas are defined).

Observation: All valuations in the sense of the previous section are propositional valuations, but not vice versa.

Definition: A propositionally valid sequent is one in which any propositional valuation which is defined on all formulas involved and sends all formulas on the left to 1 sends some formula on the right to 1. Note that all propositionally valid sequents will be valid, but not vice versa (a formula which is not propositionally valid may be valid for other logical reasons).

Observation: All the Lemmas above remain true when "valid" is replaced with "propositionally valid".

Theorem: If a sequent ϕ is propositionally valid, applications of the rules above will inevitably show this. If a sequent ϕ is not propositionally valid, applications of the rules above will inevitably reduce the sequent to a form from which a valuation witnessing its invalidity can be extracted.

Proof: Any application of the rules above converts a sequent with n disjunctions and negations in it to one or two sequents with $n-1$ disjunctions and negations each. So sufficiently many applications of the rules will convert any sequent into a collection of sequents in which all formulas are atomic (or quantified), but in any event do not have accessible disjunctions or negations. If each of these sequents has a formula in common between its left and right sets, the sequent is valid. If one of these sequents does not have a formula in common between its left and right sides, a valuation assigning 1 to each formula on the left and 0 to each formula on the right witnesses the fact that the original formula is (propositionally) invalid. The total number of steps will be no more than $1+2+4+\ldots+2^n = 2^{n+1}-1$ (which means that proofs of complex sequents may be impractically large!), because if we start with a sequent with n connectives and organize our work into steps in which we apply a single rule to each sequent, at step k we will obtain no more than 2^k formulas of length $n-k$.

So we have given a complete formal account of propositional logic.

It is worth noting that a form of the rules above can be given in which all sequents have the empty set or a singleton set on the right. Many readers will be comfortable with many premisses but only a single intended conclusion (the case of the empty set represents the goal of a contradiction). This can be done purely mechanically: apply the rules in the forms given above, then,

if there is more than one formula on the right, convert all but one of them to their negations and move them to the left. In the case of the negation rule, move the original conclusion to the left; in the case of the right rule for disjunction, move the second disjunct to the left. The theorem still holds.

The given rules can be used to derive rules for the other propositional connectives. These resemble the proof strategies that we have developed in the chapter on Proof, with the notable exception that the left rule for implication seems different (although it does support the modus ponens and modus tollens strategies we expect). The resemblance of the sequent rules to our proof strategies is clearer in the single-conclusion forms (though the left rule for implication remains eccentric).

We can present sequent proofs as mathematical objects.

Definition: An axiom (a sequent with nonempty intersection between the left and right side) is a proof of its own validity.

If the validity of sequent A follows from the validity of sequent B by an application of a sequent rule, and C is a proof of B, then $\langle A, C\rangle$ is a proof of A.

If the validity of sequent A follows from the validity of sequents B and C by an application of a sequent rule, and D is a proof of A and E is a proof of C, then $\langle A, \langle D, E\rangle\rangle$ is a proof of A.

Being an instance of one of the sequent rules is mathematically definable, so the notion of being a sequent proof is mathematically definable (the class of sequent proofs is the smallest class with the closure conditions just described).

Note that the addition of more sequent rules will cause only minor adjustments to this definition.

A sequent is *provable* if there is a proof of it. A sentence ϕ is provable iff the sequent $\vdash \phi$ is provable.

We give the propositional sequent rules in a useful format. In each entry, the validity of the sequent below the line is equivalent to all the sequents above the line being valid.

$$\Gamma, A \vdash A, \Delta$$

$$\frac{\Gamma \vdash A, \Delta}{\Gamma, \neg A \vdash \Delta}$$

$$\frac{\Gamma, A, B \vdash \Delta}{\Gamma, A \wedge B \vdash \Delta}$$

$$\frac{\Gamma, A \vdash \Delta \qquad \Gamma, B \vdash \Delta}{\Gamma, A \vee B \vdash \Delta}$$

$$\frac{\Gamma \vdash A, \Delta \qquad \Gamma, B \vdash \Delta}{\Gamma, A \to B \vdash \Delta}$$

$$\frac{\Gamma, A \to B, B \to A \vdash \Delta}{\Gamma, A \leftrightarrow B \vdash \Delta}$$

$$\frac{\Gamma, A \vdash \Delta}{\Gamma \vdash \neg A, \Delta}$$

$$\frac{\Gamma \vdash A, B, \Delta}{\Gamma \vdash A \vee B, \Delta}$$

$$\frac{\Gamma \vdash A, \Delta \qquad \Gamma \vdash B, \Delta}{\Gamma \vdash A \wedge B, \Delta}$$

$$\frac{\Gamma, A \vdash B, \Delta}{\Gamma \vdash A \to B, \Delta}$$

$$\frac{\Gamma, A \vdash B, \Delta \qquad \Gamma, B \vdash A, \Delta}{\Gamma \vdash A \leftrightarrow B, \Delta}$$

4.6 Formal First-Order Sequent Calculus: the Completeness, Compactness and Löwenheim-Skolem Theorems

For first-order reasoning, we need to introduce sequent rules for quantification.

Lemma (Cut Rule): $\Gamma \vdash \Delta$ is valid iff $\Gamma, A \vdash \Delta$ and $\Gamma \vdash A, \Delta$ are both valid.

This may seem like a purely propositional rule, though we did not need it in the previous section. As we will see in a later subsection, we do not need it here either, but it is very convenient.

We give the sequent rules for quantifiers (and the Hilbert symbol).

Lemma: $\Gamma, (\exists x.\phi[x]) \vdash \Delta$ is valid iff $\Gamma, \phi[a] \vdash \Delta$ is valid, where a is a constant which does not appear in the first sequent.

Lemma: $\Gamma \vdash (\forall x.\phi[x]), \Delta$ is valid iff $\Gamma \vdash \phi[a], \Delta$ is valid, where a is a constant which does not appear in the first sequent.

Lemma: $\Gamma \vdash (\exists x.\phi[x]), \Delta$ is valid iff $\Gamma \vdash \phi[t], (\exists x.\phi[x]), \Delta$ is valid, where t is any term.

Lemma: $\Gamma, (\forall x.\phi[x]) \vdash \Delta$ is valid iff $\Gamma, (\forall x.\phi[x]), \phi[t] \vdash \Delta$ is valid, where t is any term.

The sequent rules for equality are the following.

Lemma: Any sequent of the form $\Gamma \vdash t = t, \Delta$ is valid. We take these as axioms.

Lemma: $\Gamma, t = u \vdash \phi[t], \Delta$ is valid iff $\Gamma, t = u \vdash \phi[u], \Delta$ is valid.

Here are the rules for the Hilbert symbol.

Lemma: For any term t, $\Gamma \vdash \Delta$ is valid iff $\Gamma, \phi[(\epsilon x.\phi)/x] \vdash \Delta$ is valid and $\Gamma \vdash \phi[t/x], \Delta$ is valid. If the existential quantifier is defined in terms of the Hilbert symbol, its rule can be derived from this rule (and the rule for the universal quantifier from the rule for the existential quantifier). Note that the Cut Rule is actually a special case of this rule.

Lemma: $\Gamma, \phi[(\epsilon x.\psi[x])/x], (\forall x.\psi[x] \leftrightarrow \chi[x]) \vdash \phi[(\epsilon x.\chi[x])/x], \Delta$.

Here is a lemma about provability which follows from common features of all our rules.

Lemma: If $\Gamma \vdash \Delta$ is provable using our sequent rules then $\Gamma, \Gamma' \vdash \Delta, \Delta'$ is also provable using our sequent rules for any finite sets Γ', Δ'.

These rules correspond precisely to our proof strategies for proof of quantified goals and use of quantified hypotheses. Our definition of proofs as formal objects can be extended to first order logic proofs by adding these sequent rules.

We now prove a constellation of results which show that first order logic is *complete* (any valid sequent can be proved using the rules we have given) but also cast some doubt on just how strong first-order logic is.

Observation: The sets of formal terms and formulas are countably infinite. It is obvious that they have countably infinite subsets, so they are not finite. A quick way to see that they are just countably infinite is to observe that all our objects (formulas, terms, sequents, and proofs) are built from natural numbers by pairing and the construction of finite sets, and that finite sets and pairs of natural numbers can be implemented as natural numbers, as we showed above. So the sets of terms and formulas could be understood as infinite sets of natural numbers. The formulation we use is advantageous because it is clearly adaptable to larger languages (we might for example want uncountably many constants). This argument also adapts to larger languages: for any set of an infinite cardinality κ, objects in the set can be used to code pairs of objects in the set by the theorem $\kappa^2 = \kappa$ of cardinal arithmetic, so if we for example have κ constants and otherwise the usual finite or countable complement of symbols we will have formula and term sets of size κ.

It is also important to note that the Construction which follows is valid for restricted languages. Limiting the number of constants, predicates, functions, relations and/or operators to a finite set does not affect the construction. Completely eliminating the Hilbert symbol does not affect the Construction. Using just one type or a finite subset of the types does not affect the Construction.

Construction: Let Γ and Δ be possibly infinite sets of formulas with the property that for any finite $\Gamma_0 \subseteq \Gamma$ and $\Delta_0 \subseteq \Delta$, $\Gamma_0 \vdash \Delta_0$ is not provable, and which are such that infinitely many constants of each type do not appear in any formula in either of these sets (this is not an essential limitation: constants a_i used can be replaced with a_{2i} in each type, freeing up infinitely many constants). Then there is a countably infinite structure in which each formula in Γ and the negation of each formula in Δ is satisfied.

For purposes of this proof we use only negation, disjunction, and the existential quantifier as logical operations (all the others are definable in terms of these and their proof rules are derivable from the rules for these and their definitions).

The fact that the model constructed will be countably infinite will be evident, because the elements of the model will be terms.

We provide an enumeration F_i of all the formulas of our language in which no bound variable appears free (every bound variable is in the scope of a quantifier over that variable), and in which shorter formulas appear before longer formulas.

We define sequences of finite sets of formulas Γ_i and Δ_i which will have the following properties.

1. Each $\Gamma_i, \Gamma' \vdash \Delta_i, \Delta'$ is not a provable sequent for any finite subsets Γ', Δ' of Γ, Δ respectively.
2. $\Gamma_i \subseteq \Gamma_{i+1}$; $\Delta_i \subseteq \Delta_{i+1}$
3. Each formula F_i appears in $\Gamma_{i+1} \cup \Delta_{i+1}$

The motivation is that the set Γ_∞ which is the union of all the Γ_i's will be the set of true statements of the model to be constructed and the set Δ_∞ which is the union of all the Δ_i's will be the set of false statements of the model to be constructed.

$\Gamma_0 = \Delta_0 = \emptyset$. The conditions are clearly satisfied so far.

If Γ_i and Δ_i are defined and the conditions are supposed satisfied so far, we have next to consider where to put F_i.

1. If $\Gamma_i, \Gamma' \vdash F_i, \Delta'$ is not provable for any finite subsets Γ', Δ' of Γ, Δ respectively, set $\Gamma_{i+1} = \Gamma_i$ and $\Delta_{i+1} = \Delta_i \cup \{F_i\}$.
2. If $\Gamma_i, \Gamma' \vdash F_i, \Delta'$ is provable for some $\Gamma' \subseteq \Gamma$ and $\Delta' \subseteq \Delta$, then it cannot be the case for any finite subsets Γ'', Δ'' of Γ, Δ respectively that $\Gamma_i, \Gamma'', F_i \vdash \Delta''$ is provable, as we would then be able to prove $\Gamma', \Gamma'' \vdash \Delta', \Delta''$ using the Cut Rule. If F_i is not of the form $(\exists x.\phi[x])$, we define Γ_{i+1} as $\Gamma_i \cup \{F_i\}$ and $\Delta_{i+1} = \Delta_i$. If F_i is of the form $(\exists x.\phi[x])$, let a be the first constant of the same type as x which does not appear in any formula in Γ, Δ, Γ_i or Δ_i, let Γ_{i+1} be defined as $\Gamma_i \cup \{(\exists x.\phi[x]), \phi[a]\}$ and let Δ_{i+1} be defined as Δ_i [an important alternative is to use the Hilbert symbol $(\epsilon x.\phi[x])$ instead of a]. Note that if $\Gamma_i, (\exists x.\phi[x]), \phi[a], \Gamma' \vdash \Delta_i, \Delta'$ were provable, so would $\Gamma_i, (\exists x.\phi[x]), \Gamma' \vdash \Delta_i, \Delta'$, and we have already pointed out that the latter cannot be proved for any subsets Γ', Δ' of Γ, Δ respectively in this case. [If the alternative approach is used, note that if $\Gamma_i, (\exists x.\phi[x]), \phi[(\epsilon x.\phi[x])/x], \Gamma' \vdash \Delta_i, \Delta'$ were provable, then $\Gamma_i, (\exists x.\phi[x]), \Gamma' \vdash \Delta_i, \Delta'$ would also be provable].

The discussion shows that the conditions required continue to hold at each stage of the construction. So the definition succeeds and we obtain sets Γ_∞ and Δ_∞ whose union is the set of all formulas and whose properties we now investigate.

We are able to show that the following Lemmas hold.

Lemma: Γ_∞ and Δ_∞ are disjoint.

Proof: If they had a common element A, then some Γ_i and Δ_i would have that common element, and $\Gamma_i \vdash \Delta_i$ would be an axiom of sequent calculus.

Lemma: $\Gamma \subseteq \Gamma_\infty$; $\Delta \subseteq \Delta_\infty$

Proof: Consider what happens to F_i in either of these sets at the appropriate stage of the Construction.

Lemma: $\neg\phi \in \Gamma_\infty \leftrightarrow \phi \in \Delta_\infty$; equivalently, $\neg\phi \in \Gamma_\infty$ iff ϕ is not in Γ_∞.

Proof: Otherwise for some i, Γ_i would contain both ϕ and $\neg\phi$ or Δ_i would contain both ϕ and $\neg\phi$. In either case $\Gamma_i \vdash \Delta_i$ would be provable.

Lemma: $\phi \vee \psi \in \Gamma_\infty$ iff either $\phi \in \Gamma_\infty$ or $\psi \in \Gamma_\infty$.

Proof: Otherwise we would either have $\phi \vee \psi$ in Γ_∞ and both ϕ and ψ in Δ_∞, in which case

$$\Gamma_i \vdash \Delta_i$$

for some i would take the form $\Gamma_i, \phi \vee \psi \vdash \phi, \psi, \Delta_i$, which would be provable, or we would have $\phi \vee \psi \in \Gamma_\infty$ and either $\phi \in \Delta_\infty$ or $\psi \in \Delta_\infty$, and thus some $\Gamma_i \vdash \Delta_i$ would take one of the forms

$$\Gamma_i, \phi \vdash \phi \vee \psi, \Delta_i$$

or

$$\Gamma_i, \psi \vdash \phi \vee \psi, \Delta_i,$$

both of which are provable.

Lemma: $(\exists x.\phi[x]) \in \Gamma_\infty$ iff there is a term t such that $\phi[t] \in \Gamma_\infty$.

Proof: If $(\exists x.\phi[x]) = F_i$ and $(\exists x.\phi[x]) \in \Gamma_\infty$ then some $\phi[a]$ is also in Γ_∞ by a specific provision of the construction. If $(\exists x.\phi[x]) \in \Delta_\infty$ and there is some $\phi[t] \in \Gamma_\infty$, then some $\Gamma_i \vdash \Delta_i$ takes the form

$$\Gamma_i, \phi[t] \vdash (\exists x.\phi[x]), \Delta_i$$

and this is provable.

Lemma: $t = t \in \Gamma_\infty$ for any term t. If $t = u \in \Gamma_\infty$ and $\phi[t] \in \Gamma_\infty$ then $\phi[u] \in \Gamma_\infty$.

Proof: Immediate from the form of the sequent rules for equality.

Lemma: The relation $=_n$ on terms of type n which holds between terms t and u of type n just in case $t = u \in \Gamma_\infty$ is an equivalence relation.

Proof: $t = u \vdash u = t$ and $t = u, u = v \vdash t = v$ are provable.

Lemma: For any term t, if $\phi[t/x] \in \Gamma_\infty$ then $\phi[(\epsilon x.\phi[x])/x] \in \Gamma_\infty$.

Proof: $\phi[t/x] \vdash \phi[(\epsilon x.\phi[x])/x]$ is provable.

Lemma: If $(\forall x.\phi[x] \leftrightarrow \psi(x)) \in \Gamma_\infty$ then $(\epsilon x.\phi[x]) = (\epsilon x.\psi[x]) \in \Gamma_\infty$.

Now we can define the interpretation of our language that we want. The elements of D_n are the terms of type n in our language. a_i^n is actually defined as $a_i^{\mathbf{n}}$ (each constant is its own referent). F_i^n is the map which sends each type n term t to the term $F_i(t)$. O_i^n sends each pair of type n terms $\langle t, u\rangle$ to the term $t\, O_i\, u$. P_i^n is the set of all terms t of type n such that $P_i(t) \in \Gamma_\infty$. R_i^n is the set of all pairs of type n terms $\langle t, u\rangle$ such that $t\, R_i\, u \in \Gamma_\infty$. The functions H_n are chosen so that $H_n(\{t \mid \phi[t/x] \in \Gamma_\infty\})$ is the formal term $(\epsilon x.\phi)$.

The idea here is that we construct a model in which each term is taken to represent itself. The atomic formulas are evaluated in a way consistent with the idea that $\phi \in \Gamma_\infty$ simply means "ϕ is true in the term model", and the lemmas above show that complex terms and formulas are evaluated exactly as they should be for this to work. We conclude that for each formula $\phi \in \Gamma$, ϕ is satisfied in the term model, and for each formula $\phi \in \Delta$, ϕ is not satisfied ($\neg\phi$ is satisfied) in the term model.

Definition: For any environment E whose range consists of closed terms and term or proposition T, we define $T[E]$ as $T[E(x_1)/x_1][E(x_2)/x_2]\ldots[E(x_n)/x_n]$ where n is the largest index of a variable which occurs free in T.

Theorem: In the interpretation of our language just described, $\mathcal{V}(E, \phi) = 1 \leftrightarrow \phi[E] \in \Gamma_\infty$ for each formal sentence ϕ, and $\mathcal{R}(E, t) = t[E]$ for each formal term t.

Indication of Proof: This is proved by induction on the structure of formal terms and propositions. The Lemmas above provide the key steps.

The following theorems follow from considering the Construction and following Theorem.

Completeness Theorem: Any valid sequent has a proof.

Proof: This is equivalent to the assertion that any sequent which is not provable is invalid. A sequent $\Gamma \vdash \Delta$ is invalid precisely if there is an interpretation of the language under which Γ consists entirely of true statements and Δ consists entirely of false statements. The Construction shows us how to do this for any sequent which cannot be proved.

Definition: A collection Γ of sentences is *consistent* iff there is an interpretation under which all of them are true.

Compactness Theorem: Any collection of sentences any finite subcollection of which is consistent is consistent.

Proof: Let Γ be a collection of sentences any finite subcollection of which is consistent. This implies that $\Gamma_0 \vdash \emptyset$ is invalid for each finite $\Gamma_0 \subseteq \Gamma$. This means that $\Gamma \vdash \emptyset$ satisfies the conditions of the Construction so there is an interpretation in a term model under which all the sentences in Γ are true.

Löwenheim-Skolem Theorem: Any consistent set of sentences has a finite or countable model. If it has models of every finite size it has an infinite model.

Proof: Any consistent set of sentences satisfies the conditions of the Construction, and so has a term model, which is countable (or finite). If the theory has models of every finite size, it is consistent with the theory resulting if we adjoin new constants a_i indexed by the natural numbers with axioms $a_i \neq a_j$ for each $i \neq j$, by Compactness. A model of this extended theory will of course be infinite.

The relation $=^n$ on D_n implementing on each type n will not be the equality relation on D_n, but it will be an equivalence relation. We can convert any model in which equality is represented by a nontrivial equivalence relation into one in which the equality relation is represented by the true equality relation on each type by replacing model elements of type n with

their equivalence classes (or representatives of their equivalence classes) under $=^n$.

If the logic is extended to support our type theory, equality is definable. The relation $(\forall z.x \in z \rightarrow y \in z)$ provably has the properties of equality in the presence of the axiom of comprehension. Unfortunately, as we will see in the next section, full type theory does not satisfy the Completeness Theorem.

Although the set-theoretical definition of the Hilbert symbol involves Choice (and if we add type theory as part of our logic without some care, the properties of the Hilbert symbol will imply Choice) the Hilbert symbol adds no strength to first-order logic. If we have any theory not using the Hilbert symbol, we can use the Construction (without Hilbert symbols) to build an interpretation of the language of the theory in which all sentences are evaluated, and then (since the domain of this interpretation is countable), add the order $t \leq u$ on terms defined by "the first term equal to t in the interpretation appears no later than the first term equal to u in the interpretation in a given fixed order on terms". Then define $(\epsilon x.\phi[x])$ as the first object in this order such that ϕ. The definition of $\leq$ extends to the new Hilbert terms, and all formulas involving the defined Hilbert symbol have valuations determined in the interpretation.

The alternative version of the Construction in which existential statements are witnessed by Hilbert symbols instead of new constants has the immediate merit that one does not need infinitely many free constants and the additional merit that every object in the term model is definable from the basic concepts of the theory (in the original version of the Construction, the witnesses have an anonymous quality).

If our language is made larger by providing an uncountable collection of constants, predicates, and/or function symbols, say of uncountable size κ, the Construction still works, with the modification that "$\Gamma \vdash \Delta$ is provable" should systematically be read "for some finite $\Gamma_0 \subseteq \Gamma$ and $\Delta_0 \subseteq \Delta$ $\Gamma_0 \vdash \Delta_0$ is provable". The difficulty is that the construction will pass through stages indexed by ordinals, and once $\alpha \geq \omega$ we will have Γ_α and Δ_α infinite sets. Note that we are not talking here about modifications which would make terms or formulas of the language themselves into infinite objects (such as infinite conjunctions or disjunctions). The Compactness Theorem is thus seen to hold for languages of all sizes, and likewise the Löwenheim-Skolem Theorem can be extended to assert that any theory with infinite models has models of each infinite size κ: to ensure that there are many distinct

objects in a term model, add enough constants a_α with axioms $a_\alpha \neq a_\beta$ for each $\alpha \neq \beta$. Any finite collection of these new axioms will be consistent with any theory which has infinite models, and the Construction will give an interpretation under which all the new constants are distinct.

NOTE (everything to end of section):

Think about Omitting Types theorem here or later.

TNT is a nice exercise for this section. Also showing that type theory is distinct from Zermelo by showing that there are models of type theory with more natural numbers than types.

Section 6 is soon enough for development of the logic of the set constructor, but some allowance for the set constructor (and its type regime) should be added to syntax (which will require changes in my remarks). Add remarks about single-sorted theories being readily supported here, and more complex multi-sorted theories possible but not needed.

4.6.1 Exercises

1. Express the axioms of group theory in the language of first order logic (you do not need types and you do not need to use numerical codings). Groups are exactly models of this theory. A group is said to have *torsion* if there is an element g of the group and a natural number n such that g^n is the identity element e of the group. A group is said to be *torsion-free* if it does not have torsion. Prove that there is no formula ϕ in our formal language for group theory which is true of exactly the groups with torsion. Hint: use compactness. Suppose that ϕ is a formula which is true in every group with torsion. Consider the sentences τ_n which say "there is a g such that $g^n = e$" for each concrete natural number n. Notice (explain) that each of these sentence can be written in our formal language. Verify that the infinite set of sentences $\{\phi, \neg\tau_1, \neg\tau_2, \neg\tau_3 \ldots\}$ satisfies the conditions of the Compactness Theorem (give details). Draw the appropriate conclusion.

   Explain why this tells us that $(\exists n \in \mathcal{N}.g^n = e)$ is not equivalent to any sentence in our formal language for group theory.

2. The Löwenheim-Skolem Theorem tells us that every theory with a finite or countable language has a finite or countable model. Our untyped set theory has a countably infinite language, so has countably infinite models.

   But in untyped set theory Cantor's Theorem $|A| < |\mathcal{P}(A)|$ holds. As an exercise in porting results from type theory to set theory, write out the proof of Cantor's Theorem in untyped set theory. Hint: you do not need to make finicky use of the singleton operator in your argument.

   Finally, if A is an infinite set in a model of untyped set theory, either A is not countably infinite (in which case we have an uncountable set) or A is countably infinite and $|A| < |\mathcal{P}(A)|$, in which case $\mathcal{P}(A)$ is an uncountable set (according to the model). Yet the whole model may be countably infinite, and so certainly any infinite subsets of the model are countably infinite. Why is this not a contradiction (this argument is called *Skolem's paradox*)? Hint: I'm using what look like the same words in different senses here; explain exactly how.

4.7 Cut Elimination for First-Order Logic

4.8 Incompleteness and Undefinability of Truth

We say that a term t is closed iff all bound variables appearing in it are actually bound by some quantifier (or Hilbert symbol). A closed formula in this sense is a sentence. Each closed term t has a referent which we may write $\mathcal{R}(t)$ (the choice of environment will not affect the reference of a closed term). There are terms 't' such that $\mathcal{R}(\text{'}t\text{'}) = t$: '$t$' has as its referent the formal term t itself. There is a recursive procedure (using our definition of syntax) which would allow us to define a function which sends every formal term t to such a formal term 't'. Similarly we can define a function sending each formal sentence p (considered as a mathematical object) to a formal term 'p' such that $\mathcal{R}(\text{'}p\text{'}) = p$.

An additional convention will make this easier to see: let the operator O_1 be reserved to represent the ordered pair, and the constants a_{2n} to represent the natural numbers n. Since all terms are built from natural numbers by pairing, easy recursive definitions of 't' in terms of t and 'p' in terms of p can be given.

Now we can prove some quite surprising theorems.

Gödel's First Incompleteness Theorem: There is a sentence of our language which is true but cannot be proved.

Proof: Define a predicate G of formulas p as follows: $G(p)$ says "p is a formula with one free variable x and $p[\text{'}p\text{'}/x]$ is not provable". We have seen in the previous sections that everything here is definable. Let g represent the formula $G(p)$ as a mathematical object. $G(g)$ says that g is a formula with one free variable (it has one free variable p as you can see above) and $g[\text{'}g\text{'}/p]$ is not provable. But $g[\text{'}g\text{'}/p]$ is the statement $G(g)$ itself. If $G(g)$ is true, it cannot be proved. If $G(g)$ is false, it can be proved and is therefore true. So $G(g)$ is true but not provable.

There are some subtleties here if there are unintended objects among our proofs (we discussed this possibility for the natural numbers earlier). The sentence $G(g)$ cannot be provable, as we would then have a concrete proof whose existence falsifies what it proves. Suppose that $G(g)$ could be decided by being proved false: this would show that there is a "proof" of $G(g)$, but that might be an "unintended object" that we would never actually find.

This loophole can be closed by modifying the definition of G (a trick due to Rosser). Instead of constructing a statement which asserts its own unprovability, construct by the same technique a statement which asserts that if it is provable there is a shorter proof of its negation (a notion of numerical measure of size of proofs can readily be defined recursively). If a concrete proof of this statement were given, there would be a proof of its negation which was shorter, and so also concrete. If a concrete disproof of this statement were given, then the statement would be true (as no shorter statement could be a proof): this would make a concrete proof of the statement possible. Whether or not there are unintended "proofs" or "disproofs" of this statement, the statement must actually be undecidable.

This theorem applies not only to our type theory but also to bounded Zermelo set theory, Zermelo set theory and *ZFC* (where all our constructions can be carried out) and even to arithmetic (our whole formal development of the notion of provability can be carried out entirely in arithmetic: all we need is a notion of ordered pair definable in arithmetic, and we have shown that enough set theory can be defined in arithmetic that Kuratowski pairs of natural numbers can be coded as natural numbers. Even our semantics can be defined in arithmetic, with the stipulation that environments have to be partial functions from variables to domain elements (since they must be finite) and domains D_n need to be defined by formulas rather than given as sets.

A corollary of Gödel's First Incompleteness Theorem is

Gödel's Second Incompleteness Theorem: Our type theory (or untyped set theory, or arithmetic) cannot prove its own consistency.

Indication of Proof: The underlying idea is that to prove consistency is to prove that some statements cannot be proved. If the Rosser sentence can be proved, we can prove that all sentences can be proved (because if the Rosser sentence has a proof, so does its negation, and so does everything). So if we can prove consistency we must be able to prove that the Rosser sentence cannot be proved. But if we can prove that the Rosser sentence cannot be proved, then we can prove that the Rosser sentence is (vacuously) true (and so we have proved it contrary to hypothesis).

There are problems of level here. To actually prove that all this works requires results such as "if we can prove ϕ, then we can prove that ϕ is provable," and some other similar proofs along the same lines.

We have never found the First Incompleteness Theorem particularly surprising: there was never any reason to suppose that we could prove everything that happens to be true in mathematics. The Second Incompleteness Theorem is a bit more alarming (we cannot prove that the reasoning techniques in our working theory are free from paradox *in that theory*). The next result is quite alarming (and requires more care to understand).

Tarski's Theorem: The predicate of formulas p of the language of our type theory (or of untyped set theory, or of arithmetic) which asserts that p is true cannot be defined in the same theory.

Proof: Suppose there there is such a definable predicate `true`. Define $T(p)$ as "p is a predicate with one free variable x and $\neg$`true`$(p[\text{`}p\text{'}/x])$". Let t be the mathematical object representing $T(p)$. Then $T(t)$ asserts that $T(t)$ itself is not true. This is simply impossible. There can be no truth predicate (of formal sentences).

It is easy to misunderstand this. For any statement ϕ in our informal mathematical language (of whichever theory) we can say "ϕ is true"; this simply means ϕ and has nothing to do with Tarski's theorem. What we cannot do is define a predicate of formal mathematical objects Φ coding sentences ϕ of the language of our working theory in such a way that this predicate is true of Φ exactly if the corresponding formula ϕ is true in our theory. This is quite weird, since the missing predicate can be understood as a predicate of natural numbers (in any of these theories, if we construe the pair of the formalization of syntax as the pair definable on the natural numbers).

The reader should notice the formal analogy between these results (especially Tarski's Theorem) and Russell's paradox. Unfortunately here the self-application $p[\text{`}p\text{'}/x]$ cannot be exorcised as $x \in x$ was by our type discipline: the self-application is meaningful so something else has to give.

It is important to notice that the problem here is not that our theories are too weak. Any theory sufficiently strong in expressive power to describe provability (which amounts to having enough arithmetic) has these features. It should be noted that stronger theories can prove consistency of weaker

theories. For example, type theory does prove the consistency of arithmetic (because one can build a set model of arithmetic in type theory).

Chapter 5

Model Theory

NOTE: An earlier note said that all of this should be conducted in type theory. I'm not so certain, particularly as I approach Math 522 in fall 2017.

5.1 Ultrafilters and Ultrapowers

Definition: Let $\leq$ be a partial order. A nonempty subset F of $\texttt{fld}(\leq)$ is a *filter in* $\leq$ iff it has the properties that for every $x, y \in \texttt{fld}(\leq)$ there is some z such that $z \leq x$ and $z \leq y$ and that for every x, y if $x \in F$ and $x \leq y$ implies $y \in F$. A filter in $\leq$ is proper iff it is not the entire field of F. A filter in $\geq$ is called an *ideal in* $\leq$.

Definition: This is a maximally abstract definition of filters and ideals. For our purposes in this section, the partial order $\leq$ will always be the subset relation on $\mathcal{P}(X)$ for some fixed set X. So, for the rest of this section, a filter on X is a subset of $\mathcal{P}(X)$ which is a filter in the subset relation on $\mathcal{P}(X)$ in the sense just defined. Further, an *ultrafilter* on X is a filter U on X with the property that for each $A \subseteq X$, exactly one of A and $X - A$ belongs to U. Note that for each $x \in X$, the set $U_x = \{A \in \mathcal{P}(X) \mid x \in A\}$ is an ultrafilter on X; such ultrafilters are called *principal* ultrafilters on X. An ultrafilter on X which is not of the form U_x for any $x \in X$ is called a *nonprincipal* ultrafilter on X.

Theorem: Let X be an infinite set. Then there is a nonprincipal ultrafilter on X.

Proof: Choose a well-ordering W of $\mathcal{P}(X)$. We define the ultrafilter U_W by transfinite recursion. Suppose that we have determined for each $\beta < \alpha$ whether $W_\beta \in U_W$. We provide that $W_\alpha \in U_W$ iff $W_\alpha \cap \bigcap_{\beta \in F} W_\beta$ is an infinite set for each finite set F of ordinals less than α such that $W_\beta \in U_W$ for each $\beta \in F$. Notice that the case $F = \emptyset$ tells us that W_α is infinite.

We verify that U_W is an ultrafilter on X.

The intersection of any finite subset of U_W is an infinite set: we can see this by considering the last element of the finite set in terms of the well-ordering $\leq$ and applying the definition of U_W. A set A fails to belong to U_W exactly if there is a finite subcollection F of U_W such that the intersection of $F \cup \{A\}$ is finite: clearly if there is such a subcollection A is not in U_W, and if there is no such subcollection the recursive definition will place A in U_W.

We show that if A belongs to U_W and $A \subseteq B$, then B must belong to U_W: suppose B did not belong to U_W; it follows that there is a finite subcollection F of U_W such that the intersection of $F \cup \{B\}$ is finite, from which it follows that the intersection of $F \cup \{A\}$ is finite, from which it follows that A is not an element of U_W. We show that if A and B belong to U_W, there is $C \in U_W$ such that $C \subseteq A$ and $C \subseteq B$: a suitable C is $A \cap B$, for which it is clear that any finite subcollection F of U_W has the intersection of $F \cup \{A \cap B\}$ infinite because this is equal to the intersection of $(F \cup \{A\}) \cup \{B\}$. This verifies that U_W is a filter on X.

It cannot be the case that A and $X - A$ are both in U_W because their intersection is not infinite; nor can it be the case that both are not in U_W, because we would then have finite subsets F and G of U_W with the intersection of $F \cup \{A\}$ finite and the intersection of $G \cup \{X - A\}$ finite, so all but finitely many of the members of $\bigcap F$ would be outside A while all but finitely many of the members of $\bigcap G$ would be in A, so $\bigcap(F \cup G)$ would be finite, which is impossible. This verifies that U_W is an ultrafilter on X.

U_W is a nonprincipal ultrafilter because any principal ultrafilter U_x has a finite element $\{x\}$.

Note that the Axiom of Choice is used here (we have actually shown that there is a nonprincipal ultrafilter on X if $\mathcal{P}(X)$ can be well-

ordered). This use of choice is essential: it is consistent with the other axioms of type theory or set theory that there is no nonprincipal ultrafilter on any infinite set. It is easy to show that any ultrafilter on a finite set is principal.

Definition: Let X be an infinite set and let U be a nonprincipal ultrafilter on X. Let A be any set (not necessarily of the same type as X). Let f and g be two maps from X to A (these may be lateral!). We define $f \sim_U g$ as holding iff $\{x \mid f(x) = g(x)\} \in U$. It is easy to see that $\sim_U$ is an equivalence relation: reflexivity and symmetry are trivial, while transitivity follows from the fact that U is a filter: if $\{x \mid f(x) = g(x)\} \in U$ and $\{x \mid g(x) = h(x)\} \in U$, then $\{x \mid f(x) = g(x) \wedge g(x) = h(x)\} \in U$, being the intersection of two elements of U, and its superset $\{x \mid f(x) = h(x)\}$ is also in U. We define A^U, the *ultrapower* of A with respect to U, as the collection of equivalence classes under $\sim_U$. With each $a \in A$ we associate $a^* \in A^U$, defined as the equivalence class under $\sim_U$ of the constant function on X with value a. Note that the domain of $\sim_U$ is the collection of functions from X to A, and that we have indicated how to define this even if A and X are not of the same type.

Definition: Let X be an infinite set and let U be a nonprincipal ultrafilter on X. Let A and B be sets (not necessarily of the same type) and let R be a (possibly lateral) relation from A to B. For $[f]$ in A^U and $[g]$ in B^U, we define $[f]\, R^U\, [g]$ as holding iff $\{x \mid f(x)\, R\, g(x)\} \in U$ (it is straightforward to show that this does not depend on the choice of the representatives f and g of the elements of A^U and B^U). Note that $a^*\, R^U\, b^* \leftrightarrow a\, R\, b$.

Construction: We view A^U as a kind of extension of A, with each element a of A corresponding to the element a^* of A^U. We are going to define an extension of the language we use to talk about A to a language which talks about A^U. In fact, we are going to carry out such an extension for any collection of domains we wish to consider, all at once.

For any open sentence $\phi(x_1, \ldots, x_n)$ with no free variables other than $x_1, \ldots, x_n$, in which each $x_i \in A_i$, we define a sentence $\phi^*([f_1], \ldots, [f_n])$ for any fixed $[f_i] \in A_i^U$ as meaning $\{x \mid \phi(f_1(x), \ldots, f_n(x))\} \in U$.

If $f_i \equiv_U g_i$ for each i, then $\phi^*([f_1], \ldots, [f_n])$ asserts that $\{x \mid \phi(f_1(x), \ldots, f_n(x))\}$

is an element of U, and, because intersections of elements of U are in U, so is $\{x \mid \phi(f_1(x), \ldots, f_n(x)) \wedge f_1(x) = g_1(x) \wedge \ldots \wedge f_n(x) = g_n(x)\}$, which is a subset of $\{x \mid \phi(g_1(x), \ldots . g_n(x))\}$, so this latter set is in U, so $\phi^*([g_1], \ldots, [g_n])$. The argument is completely symmetrical that shows that $\phi^*([g_1], \ldots, [g_n])$ implies $\phi^*([f_1], \ldots, [f_n])$, so the choice of representatives in our notation for elements of A_i^U's is immaterial.

We note that if $\phi(x_1, \ldots, x_n)$ is $\neg\psi(x_1, \ldots, x_n)$, then $\phi^*([f_1], \ldots, [f_n])$ is equivalent to $\{x \mid \neg\psi(f_1(x), \ldots, f_n(x))\} \in U$, which is equivalent to $\{x \mid \psi(f_1(x), \ldots, f_n(x))\} \notin U$, because U is an ultrafilter, which is in turn equivalent to $\neg\psi^*([f_1], \ldots, [f_n])$. In other words, the meaning of negation in the translated language is what we expect.

If $\phi(x_1, \ldots, x_n)$ is $\psi(x_{s_1}, \ldots, x_{s_p}) \wedge \chi(x_{t_1}, \ldots, x_{t_q})$, then $\phi([f_1], \ldots, [f_n])$ is equivalent to $\{x \mid \psi(f_{s_1}(x), \ldots, f_{s_p}(x)) \wedge \chi(f_{t_1}(x), \ldots, f_{t_q}(x))\} \in U$, which is equivalent to $\{x \mid \psi(f_{s_1}(x), \ldots, f_{s_p}(x))\} \in U \wedge \{x \mid \chi(f_{t_1}(x), \ldots, f_{t_q}(x))\} \in U$, because subsets A and B of X both belong to U iff their intersection belongs to U, and this is in turn equivalent to $\psi^*([f_{s_1}], \ldots, [f_{s_p}]) \wedge \chi^*([f_{t_1}], \ldots, [f_{t_q}])$. In other words, the meaning of conjunction in the translated language is what we expect.

If $\phi(x_1, \ldots, x_n)$ is $(\exists y.\psi(y, x_1, \ldots, x_n))$, then $\phi^*([f_1], \ldots, [f_n])$ is equivalent to $\{x \mid (\exists y.\psi(y, f_1(x), \ldots, f_n(x))\} \in U$. If there is a g such that $\{x \mid \psi(g(x), f_1(x), \ldots, f_n(x))\} \in U$, then certainly $\{x \mid (\exists y.\psi(y, f_1(x), \ldots, f_n(x))\} \in U$, because $\{x \mid \psi(g(x), f_1(x), \ldots, f_n(x))\} \subseteq \{x \mid (\exists y.\psi(y, f_1(x), \ldots, f_n(x))\}$. Now suppose that $\{x \mid (\exists y.\psi(y, f_1(x), \ldots, f_n(x))\} \in U$. Define a function g such that for each x such that $(\exists y.\psi(y, f_1(x), \ldots, f_n(x))$ we have $\psi(g(x), f_1(x), \ldots, f_n(x))$: this is an application of the Axiom of Choice. Now we have $\{x \mid \psi(g(x), f_1(x), \ldots, f_n(x))\} = \{x \mid (\exists y.\psi(y, f_1(x), \ldots, f_n(x))\} \in U$ for this particular g. So we have shown that $\phi^*([f_1], \ldots, [f_n])$ iff there is a $[g]$ such that $\psi^*([g], [f_1], \ldots, [f_n])$. This means that the existential quantifier over any A_i in the base language translates to the existential quantifier over A_i^U in the extended language (here we moved the quantified argument into first position, but it should be clear that we do not really lose any generality by doing this).

Note that if $\phi(x_1, \ldots, x_n)$ is $\psi(a, x_1, \ldots, x_n)$, then $\phi^*([f_1], \ldots, [f_n])$ is equivalent to $\{x \mid \psi(a, f_1(x), \ldots, f_n(x))\} \in U$, which is equivalent to $\psi^*(a^*, [f_1], \ldots, [f_n])$, which indicates that constants taken from domains

A_i behave naturally in the extended language.

In the last two paragraphs, we have done manipulations on the first argument of an open sentence which can of course be done on any argument; since we can change the indexing of the arguments (and so of the domains) of a fixed open sentence it should be clear that we do not lose generality.

Note finally that if $\phi(x_1, \ldots, x_n)$ is true for any assignment of values to the x_i's from the appropriate A_i's, then $\{x \mid \phi(f_1(x), \ldots, f_n(x))\} = X \in U$ for any choice of f_i's, so $\phi^*([f_1], \ldots, [f_n])$ is always true. Translations of general truths about the A_i's hold true in the extended language over the A_i^U's.

5.2 Technical Methods for Consistency and Independence Proofs

There is a political point to be made here: all of these things can be done in type theory, quite naturally, and can thence be exported to *NFU* without reference to the usual set theory.

Writing in fall 2017 for Math 522 development (in which some or all of these topics will be covered) my thinking is that I will certainly want to do these in untyped set theory; but perhaps I should indicate the outlines of both approaches for the same reasons stated above.

5.2.1 Frankel-Mostowski Methods; The Independence of Choice

Possible Math 522 target.

5.2.2 Constructibility and the Minimal Model of Type Theory

Certainly a Math 522 target.

Build the Forster term model of type theory. Also, prove the consistency of CH and GCH (though this might get forced forward after the logic section, because there is model theory involved.).

5.2.3 Forcing and the Independence of CH

Certainly a Math 522 target.

The treatment of constructibility in the previous subsection is precisely that in the usual set theory (the fact that all the work is done in Z should make this clear. Our treatment of forcing is somewhat different from the treatement in the usual set theory: this can be seen from the fact that it handles atoms, which the usual techniques do not, and also from the fact that it *creates* atoms. The differences are technical: the basic idea is the same. What we do show by this method is that it appears that it is not necessary to do recursion along the cumulative hierarchy to do forcing (as is commonly done).

5.2.4 Generalizing the T operation

NOTE: this note might better belong somewhere else, but these considerations are needed here.

Certain collections, such as the natural numbers, are "the same" in each sufficiently high type. This is usually witnessed by a T operation. Some collections on which a T operation is defined get larger at each type; these are of less interest to us here.

T operations are defined on cardinals and on ordinals (more generally on isomorphism types) already. We point out that if we have defined T operations on sets A and B, there is a natural way to define a T operation on $\mathcal{P}(A)$ (for $a \subseteq A$, define $T^{\mathcal{P}(A)}(a)$ as T^A"a), on B^A (so that $T^{B^A}(f)(T^A(a)) = T^B(f(a))$, and on $A \times B$ (so that $T^{A\times B}(\langle a, b\rangle) = \langle T^A(a), T^B(b)\rangle$). We superscript T operations with their intended domains here for precision: we will not usually do this.

There is a uniform way to define T operations on sets with a certain kind of symmetry.

Definition: We call a bijection $f : V \to V$ a *permutation of the universe*. We use Π as a nonce notation for the set of all permutations of the universe. Define $j(f)$ so that $j(f)(x) = f$"x for all x ($j(f)$ is undefined on sets with urelements as members). Define $j^{\mathbf{n}}(f)$ in the obvious way. Further, we define the operation $j^{\mathbf{n}}(\iota)$ similarly (with due respect to the fact that ι is itself a type-raising operation, but the definition works formally). A set A is *n-symmetric* iff $j^{\mathbf{n}}(f)(A) = A$ for all permutations

of the universe f of the appropriate type. Notice that this implies that $A \in \mathcal{P}^{\mathbf{n}}(V)$. We define a T operation on n-symmetric objects A for each n:

$$T(A) = \{j^{\mathbf{n-1}}(f)(j^{\mathbf{n-1}}(\iota)(a)) \mid a \in A \wedge f \in \Pi\}.$$

Observation: The generalized T operation here would coincide with all T operations defined up to this point, if we used the Kuratowski ordered pair, or if we presumed that the type-level ordered pair coincided with the Quine ordered pair on sets and restricted all use of pairing to sets of sets (as would happen if we assumed strong extensionality). For cardinal numbers are 2-symmetric, isomorphism types are 4-symmetric if defined in terms of Kuratowski pairs and 2-symmetric if defined in terms of Quine pairs, and the definitions given above for power sets, function spaces, and cartesian products will coincide with appropriate T operations of this kind on power sets, function spaces and cartesian products (taking into account the effect on the degree of symmetry of these set constructions).

5.2.5 Forcing: Basic Definitions

We fix a definable partial order $\leq_P$ with field P which supports a T operation with the property that $T\text{“}(\leq_P) = \leq_P$ (which of course implies that $T\text{“}P = P$). This is of course a pun: what is being said is that the definition of P with all types raised by one will give the image under the T operation of the original partial order P. Such an order P will be defined and essentially "the same" structure in all types above a certain level.

The set P will be in some sense the space of "truth values" for the forcing interpretation. Each element of $\leq_P$ represents an (incomplete) "state of information"; the relation $p \leq_P q$ tells us that the state of information described by q extends the state of information described by p (the opposite convention is often used!). If neither $p \leq_P q$ nor $q \leq_P p$, the states of information described by p and q are to be understood to be incompatible.

"The objects of type n" of our forcing interpretation are relations x from $V^{\mathbf{n}}$ to P, that is, subsets of $V^{\mathbf{n}} \times P$, with the property that $\langle y, p\rangle \in$

$x \wedge q \geq_P p \rightarrow \langle y, q \rangle$. Notice that the type n objects of the forcing model are actually certain type $n+1$ objects. The type $n+1$ objects which will be interpreted as type n objects are called *names*. Those familiar with treatments of forcing in the usual set theory should notice that we are *not* requiring names to be relations from *names* to elements of P: this would introduce a recursion on the type structure, which is something always to be avoided in type theory. We will see below how difficulties which might be supposed to arise from this freedom in the construction of names are avoided.

The central definition of the forcing interpretation is the definition of a notation $p \vdash \phi$ for formulas ϕ of type theory, which is intended to tell us when a condition p gives us sufficient information to decide that an assertion ϕ is true.

The central theorem of the forcing interpretation will be that $p \vdash \phi$ is true for each axiom ϕ, that $p \vdash \phi$ can be deduced from $p \vdash \psi$ whenever ϕ can be deduced from ψ by a rule of logic. It will further be clear that we cannot prove $\neg(p \vdash \phi \wedge \neg\phi)$ (unless we can prove a contradiction in type theory itself). It is very important to notice that this is not metamathematics: $p \vdash \phi$ is not an assertion about a mathematical object 'ϕ' coding the assertion ϕ as in the development of Gödel's theorem or Tarski's theorem, and we are not building a set model of type theory (this cannot be done in type theory by those very theorems!). Of course we may associate with set models of type theory (if there are any) set models of type theory generated by applying a forcing interpretation to those set models, and this will be of some interest.

Definition: We define

$$\mathbb{N}_P = \{x \in \mathcal{P}(V \times P) \mid (\forall y.(\forall p \in P.(\forall q \geq_P p.\, \langle y, p \rangle \in x \rightarrow \langle y, q \rangle \in x)))\}$$

as the set of *P-names*. We define the notation $p \vdash \phi$ recursively. We suppose all logical operators defined in terms of $\wedge, \neg, \forall$.

negation: $p \vdash \neg\phi$ is defined as $(\forall q \geq_P p.\neg(q \vdash \phi))$. Informally, "no matter how much information we add to p, we will not verify ϕ".

conjunction: $p \vdash \phi \wedge \psi$ is defined as $(p \vdash \phi) \wedge (p \vdash \psi)$. This appears simple enough, but one should note that if one expands

out the definition of disjunction or implication in terms of the given definitions of negation and conjunction one does not get this nice distributivity.

universal quantification: $p \vdash (\forall x.\phi)$ is defined as $(\forall \mathbf{x} \in \mathbb{N}_P.p \vdash \phi[\mathbf{x}/x])$. Again, this definition looks very direct, but it is instructive to analyze the expansion of $p \vdash (\exists x.\phi[x])$.

pseudo-membership: (this will not be the interpretation of membership, for reasons that will become evident, but it makes the definition easier): for any x, y, $p \vdash x \in^* y$ iff $y \in \mathbb{N}_P \wedge (\forall q \geq_P T^{-1}(p).(\exists r \geq_P q.\langle x, r\rangle \in y))$. Note the necessity of the introduction of the T operator so that we have a well-formed assertion of type theory. Note also that x here is any object at all (of appropriate type) while y is a name of the next higher type.

Pseudo-membership does not officially appear in formulas of our language; this notation is used only in the definitions of equality and membership for the forcing interpretation.

equality: Let x and y be names. $p \vdash x = y$ is defined as

$$(\forall z.(p \vdash z \in^* x) \leftrightarrow (p \vdash z \in^* y)).$$

Names are asserted to be equal as soon as we have enough information to see that they have the same pseudo-members.

sethood: $p \vdash \mathtt{set}(x)$ is defined as

$$(\forall y.(p \vdash y \in^* x) \rightarrow y \in \mathbb{N}_P \wedge (\forall z.(p \vdash y = z) \rightarrow (p \vdash z \in^* x))).$$

p says that x is a set iff anything that p thinks is a pseudo-element of x is a name and any name that p thinks is equal to an pseudo-element of x p also thinks is an pseudo-element of x. We will see that under these conditions we can drop the "pseudo-".

membership: $p \vdash x \in y$ is defined as $(p \vdash x \in^* y) \wedge (p \vdash \mathtt{set}(y))$. The idea here is that we convert the names whose pseudo-extension does not respect equality to urelements. This is how we avoid recursion on type in our definitions (along with the fact that we use typically ambiguous partial orders on forcing conditions).

type-shifting convention: Notice that in atomic formulas we have p at the same type as the highest type of one of the arguments. Hereafter we stipulate $p \vdash \phi$ iff $T(p) \vdash \phi$; the type of p may freely be shifted. It would otherwise be difficult to type conjunctions, and it should be clear that this will introduce no conflicts.

NOTE: in the context of NF(U) this will be clear if the set P is strongly cantorian. What can be done (if anything) with cantorian partial orders needs to be cleared up [when it is cleared up the exact way we proceed here might need to be modified].

Chapter 6

Saving the Universe: Stratified Set Theories

This section concerns a class of untyped set theories which are related to type theory (as Zermelo set theory and *ZFC* also are) but in a different way. The first theory of this class was introduced by Quine in his "New foundations for mathematical logic" (1937) and so is called *NF*, which is short for "New Foundations". *NF*, as we shall see, is a very strange theory for rather unexpected reasons. We shall ignore historical precedent and start by introducing *NFU* (New Foundations with urelements), which is much more tractable. *NFU* was shown to be consistent by R. B. Jensen in 1969.

Most of the theories of this class share the perhaps alarming characteristic that they assert the existence of a universal set.

6.1 Introducing *NFU*

The starting point of the line of thought which led Quine to "New Foundations" but which will lead us first to *NFU* (due to careful planning) is an observation which we have already exploited. The types of our type theory are very similar to one another (in terms of what we can prove). We have used this observation to avoid cluttering our notation with endless type indices. We begin by carefully stating the facts already known to us (at least implicitly) about this ambiguity of type and considering some extrapolations.

6.1.1 Typical Ambiguity Examined

If we suppose that each variable x in the language of our type theory actually comes with a type index ($x^{\mathbf{n}}$ is the shape of the typical type n variable), we can define an operation on variables: if x is a variable of type n, we define x^+ as the variable of type $n+1$ obtained by incrementing the type index which x is supposed to have (though we continue our convention of not expressing it). This allows us to define an operation on formulas: if ϕ is a formula of the language of type theory, we define ϕ^+ as the result of replacing every variable x (free or bound) in ϕ with the type-incremented x^+. The same operation can be applied to terms: $\{x \mid \phi\}^+ = \{x^+ \mid \phi^+\}$, and $(\epsilon x.\phi)^+ = (\epsilon x^+.\phi^+)$.

Our first observation is that for any formula ϕ, ϕ^+ is also a formula, and for any term T, T^+ is also a formula. The converse is also true. Further, if ϕ is an axiom, ϕ^+ is also an axiom (in fact, the converse is also true). Further, if ψ can be deduced from ϕ by any logical rule, ψ^+ can also be deduced from ϕ^+, whence it follows that if ϕ is a theorem of type theory, ϕ^+ is also a theorem of type theory. In this case, the converse is not necessarily the case, though the converse does hold in *TNT*. This means that anything we know about a particular type (and a number of its successors) is also true in each higher type (and a number of its corresponding, appropriately type-shifted successors). Further, any object we can construct in type theory has a correlate constructed in the same way at each higher type. We have exploited this phenomenon, which Whitehead and Russell called "systematic ambiguity" in the more complex system of their *Principia Mathematica*, which most workers in the area of *NF* now call "typical ambiguity", and which is a rather extreme example of what computer scientists call *polymorphism*, to make it almost completely unnecessary to mention specific type indices in the typed set theory section of this book.

Quine made a daring proposal in the context of a type theory similar to ours (in fact, differing only in the assumption of strong extensionality). He suggested that it is not just the case that provable statements are the same at each type, but that the same statements are true in each type, and that the objects at the different types with correlated definitions do not merely serve as the subjects of parallel theorems but are in fact the same objects. The theory which results if this proposal is applied to our type theory is an untyped set theory, but rather different from the theory of Zermelo developed above.

In this theory we have not a universal set $V^{\mathbf{n+1}} = \{x^{\mathbf{n}} \mid x^{\mathbf{n}} = x^{\mathbf{n}}\}$ for each

n, but a single set $V = \{x \mid x = x\}$. We have already shown that it follows from the Axiom of Separation of Zermelo set theory that there can be no such set V (whence it follows that if this new theory is coherent it does not satisfy the Axiom of Separation). We do not have a $3^{\mathbf{n+1}}$ which contains all the three-element sets of type n objects, but a single object 3 which is the set of all three-element sets.

We will give the precise definition of this theory in the next section. What we will do now is prove a theorem due to Specker which will make the connections between various forms of typical ambiguity clearer. For the rest of this section, we discuss theories framed in languages in which variables are typed and which satisfy the condition that for any formula ϕ is well-formed if and only if ϕ^+ is well-formed. Further, we require that the language of the theory be closed under the basic logical operations familiar to us from above, and that whenever the rules allow us to deduce ϕ from ψ [neither formula mentioning any maximum type] we are also able to deduce ϕ^+ from ψ^+. It is required that every context in which a term can occur dictates the type of that term exactly.

We consider the following suite of axioms.

Ambiguity Scheme: For each sentence ϕ (formula with no free variables) for which ϕ^+ is well-formed, $\phi \leftrightarrow \phi^+$

With any theory T in typed language, we associate a theory T^∞ whose sentences are simply the sentences of T with all type distinctions removed. A model of T^∞, if there is one, is a model of the typed theory T in which all the types are actually the same. Notice that T^∞ is automatically the same as $(T + Amb)^\infty$, where Amb is the ambiguity scheme above, because Amb^∞ is a set of tautologies.

Note that the language of T^∞ allows things to be said which cannot be said in the typed language of T: sentences like $a \in a$ are well-formed, and a completion of a consistent T^∞ would assign truth values to such sentences.

Theorem (Specker): For any theory in typed language which is well-behaved in the ways outlined above, T^∞ is consistent iff $T + Amb$ is consistent.

Proof: It is obvious that the consistency of T^∞ implies the consistency of $T + Amb$.

Suppose that $T + Amb$ is consistent. Our goal is to show that T^∞ has a model. We first observe that this is obvious if the language of

T contains the Hilbert symbol (or any construction with equivalent logical properties). For $T + Amb$, being consistent, can be extended to a complete theory, which has a model consisting entirely of closed terms T built using the Hilbert symbol. We can then identify the term T with the term T^+ for every T. No conflict can occur: any assertion $\phi(T)$ has the same truth value as $\phi^+(T^+)$ (and these identifications and equivalences can be indefinitely iterated) [and no weird variants such as $\phi^+(T)$ are meaningful]. The truth value of $\phi(a_1, \ldots, a_n)$ with any Hilbert symbol arguments a_i however weirdly typed can be established by raising the type of ϕ sufficiently high that the types expected for its arguments are higher than the types of any of the a_i's then raising the types of the arguments a_i to the correct types, then evaluating this well-typed formula.

To complete the proof we need to show that any typed theory $T + Amb$ can be extended to include a Hilbert symbol in a way which preserves the truth of all sentences and allows Amb to be extended to the new sentences. Since $T + Amb$ is consistent, we can suppose it complete. We list all Hilbert symbols, stipulating that a Hilbert symbol must appear after any Hilbert symbol which occurs as a subterm of it in the list. We assume that before each Hilbert symbol is introduced we have a deductively closed theory which contains all instances of Amb appropriate to its language (i.e., not instances which mention Hilbert symbols not yet introduced). We introduce the Hilbert symbol $a = (\epsilon x.\chi[x])$. We then find a maximal collection of sentences $\phi[a]$ which includes $\chi[a]$, contains all type-raised copies of its elements, and is consistent. For any conjunction Φ of these sentences we have $(\exists x.\Phi^{+^i}(x))$ consistent for any i, so we can consistently add all $\phi^{+^i}[a^{+^i}]$ to our theory.

We now assume that we have a complete set of sentences $\Phi[a, a^+, \ldots, a^{+^k}]$ consistent with our theory and closed under $+$ (we have just dealt with the base case $k = 1$). We show that we can get a complete set of sentences $\phi[a, a^+, \ldots, a^{+^{k+1}}]$ consistent with our theory and closed under $+$. Suppose $\psi[a, a^+, \ldots, a^{+^{k+1}}]$ is a sentence which we wish to consider. We consider the status of sentences $(*) : (\exists x.\psi[a, a^+, \ldots, a^{+^k}, x] \wedge \Phi^+[a^+, \ldots, a^{+^k}, x])$ and $(*^\neg) : (\exists x.\neg\psi[a, a^+, \ldots, a^{+^k}, x] \wedge \Phi^+[a^+, \ldots, a^{+^k}, x])$, where the $\Phi^+[a^+, \ldots, a^{+^k}, x]$ represents type shifted versions of as large a conjunction of sentences from the complete set Φ as desired, which are already decided in our theory (because they mention blocks of k succes-

sive type-shifted versions of a). We see that if $\psi[a, a^+, \ldots, a^{+^k}, a^{+^{k+1}}]$ (resp. $\neg\psi[a, a^+, \ldots, a^{+^k}, a^{+^{k+1}}]$ is consistent with our theory then this statement must have already been decided as true (otherwise we would be able to disprove $\psi[a, a^+, \ldots, a^{+^k}, a^{+^{k+1}}]$ (resp. $\neg\psi[a, a^+, \ldots, a^{+^k}, a^{+^{k+1}}]$) from prior assumptions). This means that we can extend the sequence of k type shifted versions of a with a new term in such a way that the "type shifted sequence" starting with a^+ and extended with x has as many of the known properties of blocks of k type shifted versions of a as we want, and the sequence of $k+1$ elements satisfies ψ (resp. $\neg\psi$). These properties can include the ability (expressed in the formula $(*)$ (resp. $(*^\neg)$ above, which can be used to extend Φ) to further extend the sequence as many times as desired, while also preserving the property that blocks of $k+1$ elements of the extended sequence satisfy (type shifted versions of) ψ (resp. $\neg\psi$). Compactness then tells us that we can assume that all blocks of $k+1$ type shifted versions of a satisfy ψ (resp. $\neg\psi$). This means that we can proceed (again by compactness) to find a maximal collection of consistent sentences $\psi[a, a^+, \ldots, a^{+^k}, a^{+^{k+1}}]$ such that the closure of this set under $+$ is consistent with our previous theory. Repeating this process for all k gives us a theory with the new Hilbert symbol adjoined which extends *Amb* as desired. Repeating this process for all Hilbert symbols gives the desired extension of $T + Amb$ with Hilbert symbols, and with the scheme *Amb* extended appropriately to Hilbert symbols.

6.1.2 Definition and Consistency of *NFU*

We refer to the typed theory of sets which is our working theory as *TSTU* (excluding for the moment the axioms of Infinity, Ordered Pairs, and Choice). We refer to *TSTU* + strong extensionality as *TST*. We define *NFU* (for the moment) as $TSTU^\infty$, and define *NF* ("New Foundations") as TST^∞.

In this section we will expand a bit on how to understand the theory *NFU*, prove its consistency, and observe that the method of proof extends to a stronger theory which we will then make the referent of the name *NFU*.

NFU is an untyped set theory, like the theories of chapter 4. The axioms of *NFU* are exactly the axioms obtained from axioms of Extensionality and Comprehension of *TSTU* by disregarding all distinctions of type between the variables. Impossible axioms like $\{x \mid x \notin x\}$ do not appear as instances of Comprehension because $x \notin x$ is not the shape of any formula of the

language of $TSTU$: we drop the type distinctions, but this does not introduce identifications between variables.

We recapitulate the axioms of NFU.

Primitive notion: There is a designated object $\emptyset$ called the *empty set*.

Axiom of the empty set: $(\forall x.x \notin \emptyset)$.

Definition: We say that an object x is a *set* iff $x = \emptyset \vee (\exists y.y \in x)$. We write $\mathtt{set}(x)$ to abbreviate "x is a set" in formulas. We say that objects which are not sets are *atoms* or *urelements*.

Axiom of extensionality:

$$(\forall xy.\mathtt{set}(x) \wedge \mathtt{set}(y) \rightarrow x = y \leftrightarrow (\forall z.z \in x \leftrightarrow z \in y)),$$

In these axioms, the only changes we make are complete omission of references to types and type indices. The comprehension axiom is trickier.

***Axiom of comprehension:** For any formula $A[x]$ obtained by ignoring type distinctions in a formula of the language of type theory in which the variable y (of type one higher than x) does not appear,

$$(\exists y.(\forall x.x \in y \leftrightarrow A[x])).$$

We star this because it is not the form of the axiom we will use.

Definition: A formula ϕ of the language of set theory is said to be "stratified" iff there is a function σ (called a *stratification* of ϕ) from variables to natural numbers (or, equivalently, integers) such that for each atomic formula $x = y$ appearing in ϕ we have $\sigma(x) = \sigma(y)$ and for each atomic formula $x \in y$ appearing in ϕ we have $\sigma(x) + 1 = \sigma(y)$. Note that for a formula in equality and membership alone, to be stratified is precisely equivalent to being obtainable from a formula of the language of type theory by ignoring type distinctions

Axiom of stratified comprehension: For any stratified formula $A[x]$ in which the variable y does not appear,

$$(\exists y.(\forall x.x \in y \leftrightarrow A[x])).$$

The axiom of extensionality tells us that there is only one such object y which is a set (there may be many such objects y if $A[x]$ is not true for any x, but only one of them ($\emptyset$) will be a set). This suggests a definition:

Set builder notation: For any stratified formula $A[x]$, define $\{x \mid A[x]\}$ as the unique *set* of all x such that $A[x]$: this exists by Comprehension and is uniquely determined by Extensionality.

We show that *NFU* is consistent. We have shown above that it suffices to demonstrate that *TSTU*+Amb is consistent.

Let Σ be any finite collection of sentences of the language of *TSTU*. Let n be chosen so that Σ mentions only types $0-(n-1)$. Choose a sequence of sets X_i such that $|\mathcal{P}(X_i)| \leq |\iota\text{“}X_{i+1}|$ for each i. Choose injective maps $f_i : \mathcal{P}(X_i) \to \iota\text{“}X_{i+1}$ for each i and define relations $x \in_i y$ as $x \in X_i \wedge y \in X_{i+1} \wedge x \in f_i^{-1}(\{y\})$ (where of course this is understood to be false if $f_i^{-1}(\{y\})$ is undefined). It is easy to see that the resulting structure is a model of TSTU: the interpretation of a sentence of TSTU is obtained by replacing each type i variable with a variable restricted to X_i, and replacing each occurrence of $\in$ in an atomic formula $x \in y$ with $x \in_i y$, where i is the type of x. It should be easy to see that the interpretation of each axiom is true. Notice that this construction is carried out in our type theory, with the types of all the elements of the X_i's being the same fixed type whose identity does not matter for our purposes.

Now observe further that for any strictly increasing sequence s of natural numbers, the sequence X^s defined by $X^s_i = X_{s_i}$ determines an interpretation of TSTU in exactly the same way. We observe that the sentences Σ determine a partition of the n-element sets A of natural numbers as follows: consider a sequence s such that $s\text{“}\{0,\ldots,n-1\} = A$ and note the truth values of the sentences of Σ in the models X^s (which will be entirely determined by the first n terms of X^s). This is a partition of the n element subsets of $\mathbb{N}$ into no more than $2^{|\Sigma|}$ parts, which by Ramsey's theorem has an infinite homogeneous set H. Now consider any X^s such that $s\text{“}\mathbb{N} \subseteq H$: the interpretations of all sentences $\phi \leftrightarrow \phi^+$ for ϕ in the axiom scheme Amb will be true in such models. We have shown that every finite subset of Amb is consistent with TSTU, so by Compactness TSTU + Amb is consistent, so by Specker's theorem on ambiguity, *NFU* is consistent.

We have used more mathematical power than we need here. We have assumed in effect that $\beth_\omega$ exists (because we assume the existence of an

infinite sequence X_i). This is not strictly necessary: we can use a more refined form of Ramsey's theorem and show the existence of homogeneous sets of sufficient size in sufficiently long finite sequences of X_i's. However, we do not regard the existence of $\beth_\omega$ as a dubious assumption.

The method of proof used here extends to any extension of *TSTU* with ambiguous axioms. For example *NFU* + Infinity + Choice is shown to be consistent by this argument. Further, we can add the axiom of Ordered Pairs as well: add predicates π_1 and π_2 with the additional rules that typing for formulas $x\,\pi_i\,y$ follows the same rules as typing for formulas $x = y$ and additional axioms $(\forall x.(\exists! y.x\pi_i y))$ (each π_i is a function of universal domain) and $(\forall xy.(\exists! z.z\pi_1 x \wedge z\pi_2 y))$. These axioms hold in our working theory, and can be made to hold in the X_i's by stipulating that each X_i is infinite and providing bijections $\Pi : (X_i \times X_i) \to X_i$ for each i, and interpreting $x\pi_j y$ between type i objects as holding iff $y = \pi_j(\Pi_i(x))$.

Hereinafter we will usually mean *NFU* + Ordered Pairs + Choice when we refer to NFU.

We further note that a weaker form of stratification can be used. We say that a formula ϕ is *weakly stratified* iff the formula ϕ' is stratified which is obtained by replacing each occurrence of each variable free in ϕ with a distinct variable. Another way of putting this is that there is a function σ satisfying the conditions for a stratification, but only in atomic formulas in which both variables are bound. The reason that stratified comprehension entails weakly stratified comprehension is that the existence of each set $\{x \mid \phi\}$ is a special case of the existence of the sets $\{x \mid \phi'\}$ (existing by stratified comprehension) in which certain variables free in ϕ' (and so implicitly universally quantified in the axioms of comprehension in question) happen to take on the same values. An example: the set $\{x, \{y\}\}$ exists for each value of x, y (an instance of stratified comprehension) so the set $\{x, \{x\}\}$ exists for each x (an instance of weakly stratified comprehension).

We further note that stratification can be extended to a language with terms, if a stratification must take the same value at $(\epsilon x.\phi)$ that it does at x (the structure of ϕ then dictating type differentials between x and any parameters in the term), and noting that any term construction can be supposed implemented by a Hilbert epsilon term. This can be handled in the consistency proof by fixing choice functions to identify referents of Hilbert epsilon terms in the X_i's.

This proof allows us to bootstrap our working theory from *TSTU* with Ordered Pairs and Choice to *NFU* with Ordered Pairs and Choice, if we are

so inclined: we can adopt the view that the types of our theory, which are suspiciously similar because we have been careful to keep our methods of proof over them entirely uniform, are in fact all the same domain. We will explore the consequences of taking this perhaps odd view.

(NOTE: we certainly want to consider the Boffa model construction as well. For this we need enough model theory to get models with automorphisms.)

6.1.3 Mathematics in *NFU*

(NOTE: Counting is so useful that it might show up in the base development.)

We do not start with a clean slate when we consider doing mathematics in *NFU*, because all the mathematics we have done in *TSTU* can be imported. However, the interpretation of *NFU* is different in interesting ways.

The language of *NFU* is larger. Sentences such as $x \in x$ are well-formed as they are not in typed language. Further, a sentence like $V \in V$ which we wrote but construed as a sort of pun in typed language is to be taken seriously in *NFU*: the universal set V has *everything* as an element, including itself. From this it follows that $(\exists x.x \in x)$ is a theorem of *NFU*, since the universal set is a witness.

We have proved Cantor's theorem $|\iota``A| < |\mathcal{P}(A)|$ which tells us that the power set of A is larger than A. But in *NFU* we of course know that $\mathcal{P}(V) \subseteq V$. This does not contradict anything we proved in type theory, because in type theory the referents of the two V's are not supposed to be the same. In *NFU* Cantor's Theorem tells us that $|\iota``V| < |\mathcal{P}(V)| \leq |V|$, so we see that the singleton map (which from an external standpoint we can see is a one-to-one correspondence) cannot be a set in *NFU*.

The unstratified form of Cantor's Theorem which is true in the untyped set theories of chapter 4 cannot hold in general in *NFU*, but it can hold under special circumstances.

Definition: A set A is said to be *cantorian* iff $|A| = |\iota``A|$.

This is precisely what is needed to get the unstratified theorem "if A is a cantorian set, $|A| = |\iota``A| < |\mathcal{P}(A)|$". We see that all cantorian sets are smaller than their power sets. Consideration of how this fact is witnessed suggests a stronger property.

Definition: A set A is said to be *strongly cantorian* iff $(\iota\lceil A) = \{(a, \{a\}) \mid a \in A\}$ is a set.

Obviously a strongly cantorian set is cantorian. The stronger property has considerably stronger consequences.

What all of this already tells us is that a model of *NFU* is not a model of *TSTU* of the natural kind in which every collection of type i objects is a type $i+1$ object. Every element of the non-function $\iota = \{(x, \{x\}) \mid x \in V\}$ is an object in our model of *NFU*, but the collection of all these pairs cannot be an element of the model on pain of contradiction.

We give a much sharper result of the same kind. We proved above that $T^2(\Omega) < \Omega$ (recall that Ω is the order type of the ordinals). In *TSTU* this assertion was a kind of pun, but here all references to Ω are references to the same object. It is straightforward to prove that $\alpha < \beta \leftrightarrow T(\alpha) < T(\beta)$, from which it follows that $\Omega > T^2(\Omega) > T^4(\Omega) > T^6(\Omega) > \ldots$. This observation has two different rather alarming consequences. One is that a certain *countable* collection of objects of a model of *NFU* cannot be a set: if the smallest collection containing Ω and closed under T^2 were a set, it would be a set of ordinals with no smallest element, which is impossible. The other is that from a certain external standpoint, the ordinals of a model of *NFU* are not well-ordered.

We investigate the mathematics of the properties "cantorian" and "strongly cantorian".

Theorem: Concrete finite sets are cantorian. Power sets of cantorian sets are cantorian. Cartesian products of cantorian sets are cantorian. Function spaces from cantorian sets to cantorian sets are cantorian.

Proof: Sets of concrete finite sizes are obviously the same size as their images under the singleton operation. We will find that asserting this for all finite sets is a stronger assertion than we can prove from our current axioms. The other assertions follow from the existence of bijections between $\mathcal{P}(\iota"A)$ and $\iota"(\mathcal{P}(A))$, between $\iota"A \times \iota"B$ and $\iota"(A \times B)$ and between $\iota"B^{\iota"A}$ and $\iota"(B^A)$: from the ability to define these maps it clearly follows that if A, B are the same size as $\iota"A$, $\iota"B$, respectively, then $\mathcal{P}(A)$, $A \times B$, B^A are the same size as $\iota"\mathcal{P}(A)$, $\iota"(A \times B)$, $\iota"(B^A)$, respectively, which is what is to be shown.

Theorem: Concrete finite sets are strongly cantorian. Power sets of cantorian sets are strongly cantorian. Cartesian products of cantorian sets are strongly cantorian. Function spaces from cantorian sets to cantorian sets are strongly cantorian.

Proof: If A is a concrete finite set, $(\iota\lceil A)$ can be given as a concrete finite set. Again, showing that this is true for all finite sets turns out not to be provable with our current axioms. Construct $(\iota\lceil\mathcal{P}(A))$ as $(B : \mathcal{P}(A) \mapsto (A : \mathcal{P}(\iota\text{“}V) \mapsto \{\bigcup A\})((\iota\lceil A)\text{“}B))$. Construct $(\iota\lceil(A \times B))$ as $((a, b) : A \times B \mapsto ((\{x\}, \{y\}) : (\iota\text{“}V) \times (\iota\text{“}V) \mapsto \{(x, y\})((\iota\lceil A)(a), (\iota\lceil B)(b)))$. We leave the similar construction of $(\iota\lceil B^A)$ as an exercise.

Theorem: A subset of a strongly cantorian set is strongly cantorian.

Proof: If $B \subseteq A$, $(\iota\lceil B) = (\iota\lceil A)\lceil B$.

The last theorem is one reason why "strongly cantorian" is a much stronger property. Here is a further, more profound reason.

Subversion Theorem: Let ϕ be a formula in which some quantified variables are restricted to strongly cantorian sets. Let ϕ' be the formula obtained by replacing each occurrence of each variable bounded in a strongly cantorian set A with a distinct variable bounded in A (replacing single universal quantifiers over A with blocks of universal quantifiers over A or single existential quantifiers over A with blocks of existential quantifiers over A as needed). If ϕ' is stratified then $\{x \mid \phi\}$ exists. Equivalently, if there is a function which meets the conditions to be a stratification of ϕ in each atomic subformula containing two bound variables neither of which is bounded in A, then $\{x \mid \phi\}$ exists.

Proof: The formula ϕ' can be modified in such a way as to change the value assigned to a variable a restricted to the strongly cantorian set A freely. Let ι_A represent the singleton map restricted to A, for each of the strongly cantorian sets A appearing as bounds of quantifiers in ϕ. To raise the type assigned to a by one, replace a with the term $(\epsilon x.x \in \iota_A(a))$. To lower the type assigned to a by one, replace a with $\iota_A^{-1}(\{a\})$. Now each variable in ϕ' which is bounded in a strongly cantorian set A can be assigned a type in such a way that the desired additional equations between variables needed to give equivalence with ϕ can be adjoined while preserving stratification.

NOTE: other specifically NFU mathematics include unstratified inductive definitions (von Neumann ordinals, notions of well-foundedness, etc.) and T-sequences and related ideas.

6.1.4 There are Urelements

6.2 Extensions of *NFU*

6.2.1 The Axiom of Counting; ω-Models.

But perhaps Counting will be covered in the first part?

unstratified induction? The ω-model construction; α-models; NFU*.

6.2.2 The Axiom of Cantorian Sets; the Axiom of Large Ordinals

this will provide an occasion for T-sequences. Interpretation of *ZFC* in this theory (cute eliminations of T). n-Mahlos, fancy partition relations, model theory.

6.2.3 The Axiom of Small Ordinals; the BEST model

ASO with and without CS and Large Ordinals. weakly compact; nearly measurable. Solovay stuff. The BEST model.

6.3 The Extensional Subsystems

6.3.1 Ambiguity in Theories with Finitely Many Types; NF_3

Our type theory *TSTU* has natural subtheories defined simply by restricting the number of types. Similar considerations apply to variants of our type theory.

Definition: $TSTU_n$ is defined as the subtheory of *TSTU* with type indices $\geq n$ excluded from the language. Other type theories will have subscripted variants defined in the same way.

The situation in three types is very special.

Theorem: For any infinite model of $TSTU_3$ with either the same concrete finite number of atoms at each type or infinitely many atoms at each type, there is a model of $TSTU_3{}^{\infty}$ with exactly the same theory.

Proof: By model theory, there is a countable model of $TSTU_3$ with the same theory. We want a further refinement: we want a countable model with the property that each infinite set can be partitioned into two infinite sets. Suppose our initial countable model lacks this property: there are then infinite sets which can only be partitioned into finite and cofinite pieces. Construct an ultrapower of the model using an ultrafilter on the natural numbers. This will give a model of the theory with the splitting property (but not a countable one). Build a countable model with the same theory as this model, but being sure to include some specific constant (referring to a set of nonstandard finite size) in your theory. The resulting model will be countable, will have the splitting property (because we will have partitions of any infinite set with one partition of the fixed nonstandard size), and will have exactly the same theory as the original model (if we exclude references to the special constant from our language).

Now we show that in any countable model of $TSTU$ there is an isomorphism between types $0-1$ and types $1-2$. First of all, the conditions in the statement of the theorem combined with the countability of the model are enough to ensure that we have a bijection from the type 1 atoms onto the type 2 atoms. Now we handle the sets. We fix an order on the type 1 sets and an order on the type 2 sets, each of type ω. When we have mapped the first n sets of type 1 to sets of type 2, and also the first n sets of type 2 have been assigned inverse images in type 1, we assume that we have matched them in such a way that the sizes of the corresponding compartments in Venn diagrams determined by the type 1 sets assigned images and the type 2 sets assigned inverse images is correct: for any intersection of the type 1 sets and their complements, if the intersection is of concrete finite size n the corresponding intersection of type 2 sets and their complements will be of the same concrete finite size n, and if the intersection is (countably) infinite the corresponding intersection of the type 2 sets and their complements will be countably infinite. We show how to continue this process (note that the

conditions are vacuously satisfied initially). Match the first set of type 1 not yet assigned an image with the first set in the order on type 2 sets which has not yet been matched and has the correct intersection sizes with the correlates of all finite intersections of the previously mapped type 1 sets. The splitting property is needed here to ensure that if the new type 1 set has infinite and co-infinite intersection with one of the compartments of the Venn diagram determined by the previous set that we can choose a type 2 set with appropriate intersection sizes to associate with it. Choose an inverse image for the first type 2 set as yet not assigned an inverse image in exactly the same way. Notice that the map between types 1 and 2 determines a map between types 0 and 1 by considering singletons. Note that the amount of comprehension needed in the type theory considered is very limited: all that is needed is existence of singletons, complements and finite unions.

If f is the isomorphism, we take type 0 as the model and define $x \in_N y$ as $x \in_M f(y)$ (where $\in_M$ is the membership relation of the model. Note that for any $x^{\mathbf{0}} \in_M y^{\mathbf{1}}$ we have $x^{\mathbf{0}} \in_M f^{-1}(y^{\mathbf{1}})$ equivalent and for any $x^{\mathbf{1}} \in_M y^{\mathbf{2}}$ we have $f^{-1}(x^{\mathbf{1}}) \in_M f^{-2}(y^{\mathbf{2}})$ This model N will be a model of $TSTU_3{}^{\infty}$: this should be evident.

It should be evident from these considerations that all models of $TSTU_3$ satisfying the conditions on numbers of atoms (which are describable in terms of sets of sentences satisfied in their theories) also satisfy Amb (noting that the scheme $\phi \leftrightarrow \phi^+$ must be restricted to formulas not mentioning type 2).

Definition: Define NF_3 as the theory whose axioms are Strong Extensionality and those instances "$\{x \mid \phi\}$ exists" of Stratified Comprehension which can be stratified using a stratification with range $\{0, 1, 2\}$ (note that the stratification will send x to 0 or 1, since it must assign 1 or 2 to $\{x \mid \phi\}$).

Corollary: NF_3 is consistent.

Proof: In the previous Theorem, fix the number of atoms at 0.

Observation: This is the first consistent fragment of New Foundations which we have identified which has strong extensionality. It is important to notice that, unlike NF, this is *not* a weird theory involving

considerations strange to ordinary mathematics. *Every* infinite model of TST_3 has a correlated model of NF_3 which satisfies the same sentences when types are dropped. NF_3, though it may seem unfamiliar, is ubiquitous and should be of considerable interest in foundations of mathematics.

We go on to consider Ambiguity for $TSTU_n$ with $n > 3$.

Theorem: $TSTU_n^\infty$ is consistent iff $TSTU_n + Amb$ is consistent.

Proof: Notice that our proof above depended on being able to iterate the $+$ operation as far as wanted; this is spoiled by the presence of a top type. We will fix this problem using a trick.

We can cleverly delete all reference to the bottom type of our language. We define $[\subseteq]^2$ as the collection of all sets $\{x \mid x \subseteq A\}$ where A is a fixed type 1 set (it is important to recall that an urelement is not a subset of anything). We define $1^{\mathbf{1}}$ as usual as the set of all singletons. We now observe that $x^{\mathbf{0}} \in y^{\mathbf{1}}$ is equivalent to the assertion that $\{x\} \subseteq y$, which is in turn equivalent to "$\{x\}$ belongs to every element of $[\subseteq]^{\mathbf{2}}$ which contains y". We can now replace all references to specific type 0 objects by references to singletons and all quantifiers over type 0 with quantifiers over $1^{\mathbf{1}}$, redefining membership in type 0 objects appropriately.

This doesn't give us anything obvious for free, as we have our special constants 1^1 and $[\subseteq]^2$ to consider. We further observe that it is a theorem that $1^{\mathbf{1}}$ is a subset of the domain of $[\subseteq]^{\mathbf{2}}$ and for every (type 2) subset A of $1^{\mathbf{1}}$ there is a unique type 1 object a in the range of $[\subseteq]^{\mathbf{2}}$ such that "$\{x\} \subseteq a$" (a fact expressible without mentioning type 0) iff $\{x\} \in A$.

Now Amb tells us that there are objects $1^{\mathbf{0}}$ and $[\subseteq]^{\mathbf{1}}$ with the type-shifted version of the same property noted above for 1^1 and $[\subseteq]^1$. These can be reinterpreted as the singleton set on a new type -1 and the inclusion relation on type 0 objects construed as "sets" of type -1 objects. This means that $TSTU_n + Amb$ interprets $TSTU_{n+1}$ (we can reindex so that the new type -1 becomes type 0). We can further use ambiguity to ensure that as much as we wish to be true about $1^{\mathbf{0}}$ and $[\subseteq]^{\mathbf{1}}$ is the type-shifted analogue of what is true about $1^{\mathbf{1}}$ and $[\subseteq]^{\mathbf{2}}$

> [we cannot show that there are specific relations which have exactly the same properties, merely that there is a relation with any finite selection of type shifted versions of the properties of $1^{\mathbf{1}}$ and $[\subseteq]^{\mathbf{2}}$], and thus show by compactness that the extension of Amb can consistently hold as well. So the consistency of $TSTU_n + Amb$ for $n > 3$ implies the consistency of $TSTU_{n+1} + Amb$, whence it implies the consistency of $TSTU + Amb$, whence it implies the consistency of New Foundations.

Corollary: $TST_4 + Amb$ is consistent iff NF is consistent.

Observation: The profound difference between the case $n = 3$ and the case $n = 4$ in the strongly extensional case is of interest here.

Observation: The proofs above will also work in some other type theories.

Make the point that NF style considerations are natural and ubiquitous in 3-typed mathematics.

This should include the proof of consistency of NFU using 3-type machinery and the Pigeonhole Principle instead of Ramsey's theorem.

Mathematics in three types, functions without pairs. FM methods in the first section would avoid an inversion here.

6.3.2 Predicativity; NFP; The Ramified Theory of Types Interpreted in NFP; NFI

6.4 Finite Universes: *NFU* + "the universe is finite".

also NFU and nonstandard analysis?

6.5 New Foundations

6.5.1 History of *NF*; Errors of Quine

Specker trees, all the bad stuff. A section on FM methods in type theory would help here as it would provide an occasion in the first part to carefully discuss choice-free mathematics. Orey's metamathematical results; of course they also work in NFU.

6.6 Technical Methods for Consistency and Independence Proofs in *NF(U)*

6.6.1 Forcing in Type Theory and Set Theory

Introduce the method of forcing in *NFU* at least and possibly in type theory and ordinary set theory. Prove the independence of the continuum hypothesis. Forcing in *NF*, of course. But this may continue a section on forcing in the type theory part.

6.6.2 Frankel-Mostowski Permutation Methods

Prove the independence of the Axiom of Choice from type theory (certainly) and possibly from *NFU* and/or ordinary set theory. The initial parts of this may occur in the type theory part.

6.7 Cut Elimination in Type Theory and Set Theory

Prove cut elimination in type theory and *SF*. Maybe other applications of Marcel's weak extensional collapse.

6.8 Stratified Combinatory Logic and λ-Calculus

6.9 Rieger-Bernays Permutation Methods

Explore the consistency and independence proofs obtainable, and the set based notion of "well-foundedness" and related ideas. Unstratified implementations of numerals.

6.10 Limitations of Universal Constructions

The existence of universal objects is not magic. Cartesian closedness failing for the category of all sets and functions is an advantage.

Chapter 7

Philosophy of Set Theory

General considerations about the relative merits of the various systems considered here and about the sufficiency of each as a foundational system. Comments on the general weirdness of *NF* and the real nature of the *NF* consistency problem belong here.

Chapter 8

Appendix: Manual of Logical Style

This is a handout I give students at various levels with logical rules in it in the same style as the text.

8.1 Introduction

This document is designed to assist students in planning proofs. I will try to make it as nontechnical as I can.

There are two roles that statements can have in a proof: a statement can be a claim or goal, something that we are trying to prove; a statement can be something that we have proved or which we have shown to follow from current assumptions, that is, a statement which we can *use* in the current argument.[1] It is very important not to confuse statements in these two roles: this can lead to the fallacy of assuming what you are trying to prove (which is well-known) or to the converse problem, which I *have* encountered now and then, of students trying to prove things that they already know or are entitled to assume!

In the system of reasoning I present here, we classify statements by their top-level logical operation: for each statement with a particular top-level operation, there will be a rule or rules to handle goals or claims of that form,

[1] which we called a *posit* in chapter 2: I have not used this term in other contexts where I have distributed this style manual.

and a rule or rules to handle *using* statements of that form which we have proved or are entitled to assume.

In what follows, I make a lot of use of statements like "you are entitled to assume A". Notice that if you can flat-out *prove* A you are entitled to assume A. The reason I often talk about being entitled to assume A rather than having proved A is that one is often proving things using assumptions which are made for the sake of argument.

8.2 Conjunction

In this section we give rules for handling "and". These are so simple that we barely notice that they exist!

8.2.1 Proving a conjunction

To prove a statement of the form $A \wedge B$, first prove A, then prove B.

This strategy can actually be presented as a rule of inference:

$$\frac{\begin{array}{c}A\\B\end{array}}{A \wedge B}$$

If we have hypotheses A and B, we can draw the conclusion $A \wedge B$: so a strategy for proving $A \wedge B$ is to first prove A then prove B. This gives a proof in two parts, but notice that there are no assumptions being introduced in the two parts: they are not separate cases.

If we give this rule a name at all, we call it "conjunction".

8.2.2 Using a conjunction

If we are entitled to assume $A \wedge B$, we are further entitled to assume A and B. This can be summarized in two rules of inference:

$$\frac{A \wedge B}{A}$$

$$\frac{A \wedge B}{B}$$

This has the same flavor as the rule for proving a conjunction: a conjunction just breaks apart into its component parts.

If we give this rule a name at all, we call it "simplification".

8.3 Implication

In this section we give rules for implication. There is a single basic rule for implication in each subsection, and then some derived rules which also involve negation, based on the equivalence of an implication with its contrapositive. These are called derived rules because they can actually be justified in terms of the basic rules. We like the derived rules, though, because they allow us to write proofs more compactly.

8.3.1 Proving an implication

The basic strategy for proving an implication: To prove $A \to B$, add A to your list of assumptions and prove B; if you can do this, $A \to B$ follows without the additional assumption.

Stylistically, we indent the part of the proof consisting of statements depending on the additional assumption A: once we are done proving B under the assumption and thus proving $A \to B$ without the assumption, we discard the assumption and thus no longer regard the indented group of lines as proved.

This rule is called "deduction".

The indirect strategy for proving an implication: To prove $A \to B$, add $\neg B$ as a new assumption and prove $\neg A$: if you can do this, $A \to B$ follows without the additional assumption. Notice that this amounts to proving $\neg B \to \neg A$ using the basic strategy, which is why it works.

This rule is called "proof by contrapositive" or "indirect proof".

8.3.2 Using an implication

modus ponens: If you are entitled to assume A and you are entitled to assume $A \to B$, then you are also entitled to assume B. This can be written as a rule of inference:

$$\frac{\begin{array}{l}A\\A \to B\end{array}}{B}$$

when you just have an implication: If you are entitled to assume $A \to B$, you may at any time adopt A as a new goal, for the sake of proving B, and as soon as you have proved it, you also are entitled to assume B. Notice that no assumptions are introduced by this strategy. This proof strategy is just a restatement of the rule of *modus ponens* which can be used to suggest the way to proceed when we have an implication without its hypothesis.

modus tollens: If you are entitled to assume $\neg B$ and you are entitled to assume $A \to B$, then you are also entitled to assume $\neg A$. This can be written as a rule of inference:

$$\frac{\begin{array}{l}A \to B\\\neg B\end{array}}{\neg A}$$

Notice that if we replace $A \to B$ with the equivalent contrapositive $\neg B \to \neg A$, then this becomes an example of modus ponens. This is why it works.

when you just have an implication: If you are entitled to assume $A \to B$, you may at any time adopt $\neg B$ as a new goal, for the sake of proving $\neg A$, and as soon as you have proved it, you also are entitled to assume $\neg A$. Notice that no assumptions are introduced by this strategy. This proof strategy is just a restatement of the rule of *modus tollens* which can be used to suggest the way to proceed when we have an implication without its hypothesis.

8.4 Absurdity

The symbol $\perp$ represents a convenient fixed false statement. The point of having this symbol is that it makes the rules for negation much cleaner.

8.4.1 Proving the absurd

We certainly hope we never do this except under assumptions! If we are entitled to assume A and we are entitled to assume $\neg A$, then we are entitled to assume $\bot$. Oops! This rule is called *contradiction*.

$$\frac{A \qquad \neg A}{\bot}$$

8.4.2 Using the absurd

We hope we never really get to use it, but it is very useful. If we are entitled to assume $\bot$, we are further entitled to assume A (no matter what A is). From a false statement, anything follows. We can see that this is valid by considering the truth table for implication.

This rule is called "absurdity elimination".

8.5 Negation

The rules involving just negation are stated here. We have already seen derived rules of implication using negation, and we will see derived rules of disjunction using negation below.

8.5.1 Proving a negation

direct proof of a negation (basic): To prove $\neg A$, add A as an assumption and prove $\bot$. If you complete this proof of $\bot$ with the additional assumption, you are entitled to conclude $\neg A$ without the additional assumption (which of course you now want to drop like a hot potato!). This is the direct proof of a negative statement: proof by contradiction, which we describe next, is subtly different.

Call this rule "negation introduction".

proof by contradiction (derived): To prove a statement A of any logical form at all, assume $\neg A$ and prove $\bot$. If you can prove this under the additional assumption, then you can conclude A under no additional assumptions. Notice that the proof by contradiction of A is a direct

proof of the statement $\neg\neg A$, which we know is logically equivalent to A; this is why this strategy works.

Call this rule "reductio ad absurdum".

8.5.2 Using a negation:

double negation (basic): If you are entitled to assume $\neg\neg A$, you are entitled to assume A. Call this rule "double negation elimination".

contradiction (basic): This is the same as the rule of contradiction stated above under proving the absurd: if you are entitled to assume A and you are entitled to assume $\neg A$, you are also entitled to assume $\bot$. You also feel deeply queasy.

$$\frac{A \quad \neg A}{\bot}$$

if you have just a negation: If you are entitled to assume $\neg A$, consider adopting A as a new goal: the point of this is that from $\neg A$ and A you would then be able to deduce $\bot$ from which you could further deduce whatever goal C you are currently working on. This is especially appealing as soon as the current goal to be proved becomes $\bot$, as the rule of contradiction is the only way there is to prove $\bot$.

8.6 Disjunction

In this section, we give basic rules for disjunction which do not involve negation, and derived rules which do. The derived rules can be said to be the default strategies for proving a disjunction, but they *can* be justified using the seemingly very weak basic rules (which are also very important rules, but often used in a "forward" way as rules of inference). The basic strategy for using an implication (proof by cases) is of course very often used and very important. The derived rules in this section are justified by the logical equivalence of $P \vee Q$ with both $\neg P \to Q$ and $\neg Q \to P$: if they look to you like rules of implication, that is because somewhere underneath they are.

8.6.1 Proving a disjunction

the basic rule for proving a disjunction (two forms): To prove $A \vee B$, prove A. Alternatively, to prove $A \vee B$, prove B. You do *not* need to prove both (you should not expect to be able to!)

This can also be presented as a rule of inference, called *addition*, which comes in two different versions.

$$\frac{A}{A \vee B}$$

$$\frac{B}{A \vee B}$$

the default rule for proving a disjunction (derived, two forms): To prove $A \vee B$, assume $\neg B$ and attempt to prove A. If A follows with the additional assumption, $A \vee B$ follows without it.

Alternatively (do not do both!): To prove $A \vee B$, assume $\neg A$ and attempt to prove B. If B follows with the additional assumption, $A \vee B$ follows without it.

Notice that the proofs obtained by these two methods are proofs of $\neg B \to A$ and $\neg A \to B$ respectively, and both of these are logically equivalent to $A \vee B$. This is why the rule works. Showing that this rule can be derived from the basic rules for disjunction is moderately hard.

Call both of these rules "disjunction introduction", or "alternative elimination".

8.6.2 Using a disjunction

proof by cases (basic): If you are entitled to assume $A \vee B$ and you are trying to prove C, first assume A and prove C (case 1); then assume B and attempt to prove C (case 2).

Notice that the two parts are proofs of $A \to C$ and $B \to C$, and notice that $(A \to C) \wedge (B \to C)$ is logically equivalent to $(A \vee B) \to C$ (this can be verified using a truth table).

This strategy is very important in practice.

disjunctive syllogism (derived, various forms): If you are entitled to assume $A \vee B$ and you are also entitled to assume $\neg B$, you are further entitled to assume A. Notice that replacing $A \vee B$ with the equivalent $\neg B \to A$ turns this into an example of modus ponens.

If you are entitled to assume $A \vee B$ and you are also entitled to assume $\neg A$, you are further entitled to assume B. Notice that replacing $A \vee B$ with the equivalent $\neg A \to B$ turns this into an example of modus ponens.

Combining this with double negation gives further forms: from B and $A \vee \neg B$ deduce A, for example.

Disjunctive syllogism in rule format:

$$\frac{\begin{array}{r} A \vee B \\ \neg B \end{array}}{A}$$

$$\frac{\begin{array}{r} A \vee B \\ \neg A \end{array}}{B}$$

Some other closely related forms which we also call "disjunctive syllogism":

$$\frac{\begin{array}{r} A \vee \neg B \\ B \end{array}}{A}$$

$$\frac{\begin{array}{r} \neg A \vee B \\ A \end{array}}{B}$$

8.7 Biconditional

Some of the rules for the biconditional are derived from the definition of $A \leftrightarrow B$ as $(A \to B) \wedge (B \to A)$. There is a further very powerful rule allowing us to use biconditionals to justify replacements of one expression by another.

8.7.1 Proving biconditionals

the basic strategy for proving a biconditional: To prove $A \leftrightarrow B$, first assume A and prove B; then (finished with the first assumption) assume B and prove A. Notice that the first part is a proof of $A \rightarrow B$ and the second part is a proof of $B \rightarrow A$.

Call this rule "biconditional deduction".

derived forms: Replace one or both of the component proofs of implications with the contrapositive forms. For example one could first assume A and prove B, then assume $\neg A$ and prove $\neg B$ (changing part 2 to the contrapositive form).

8.7.2 Using biconditionals

The rules are all variations of modus ponens and modus tollens. Call them biconditional modus ponens (bimp) or biconditional modus tollens (bimt) as appropriate.

If you are entitled to assume A and $A \leftrightarrow B$, you are entitled to assume B.

If you are entitled to assume B and $A \leftrightarrow B$, you are entitled to assume A.

If you are entitled to assume $\neg A$ and $A \leftrightarrow B$, you are entitled to assume $\neg B$.

If you are entitled to assume $\neg B$ and $A \leftrightarrow B$, you are entitled to assume $\neg A$.

These all follow quite directly using modus ponens and modus tollens and one of these rules:

If you are entitled to assume $A \leftrightarrow B$, you are entitled to assume $A \rightarrow B$.

If you are entitled to assume $A \leftrightarrow B$, you are entitled to assume $B \rightarrow A$.

The validity of these rules is evident from the definition of a biconditional as a conjunction.

8.8 Calculating with biconditionals

Let F be a complex expression including a propositional letter P. For any complex expression C let $F[C/P]$ denote the result of replacing all occurrences of P by C.

The replacement rule for biconditionals says that if you are entitled to assume $A \leftrightarrow B$ and also entitled to assume $F[A/P]$, then you are entitled to assume $F[B/P]$. Also, if you are entitled to assume $A \leftrightarrow B$ and also entitled to assume $F[B/P]$, then you are entitled to assume $F[A/P]$.

The underlying idea which we here state very carefully is that $A \leftrightarrow B$ justifies substitutions of A for B and of B for A in complex expressions. This is justified by the fact that all our operations on statements depend only on their truth value, and $A \leftrightarrow B$ is equivalent to the assertion that A and B have the same truth value.

This rule and a list of biconditionals which are tautologies motivates the "boolean algebra" approach to logic.

8.9 Universal Quantifier

This section presents rules for $(\forall x.P(x))$ ("for all x, $P(x)$") and for the restricted form $(\forall x \in A.P(x))$ ("for all x in the set A, $P(x)$"). Notice that $(\forall x \in A.P(x))$ has just the rules one would expect from its logical equivalence to $(\forall x.x \in A \rightarrow P(x))$.

8.9.1 Proving Universally Quantified Statements

To prove $(\forall x.P(x))$, first introduce a name a for a completely arbitrary object. This is signalled by a line "Let a be chosen arbitrarily". This name should not appear in any earlier lines of the proof that one is allowed to use. The goal is then to prove $P(a)$. Once the proof of $P(a)$ is complete, one has proved $(\forall x.P(x))$ and should regard the block beginning with the introduction of the arbitrary name a as closed off (as if "Let a be arbitrary" were an assumption). The reason for this is stylistic: one should free up the use of the name a for other similar purposes later in the proof.

To prove $(\forall x \in A.P(x)$, assume $a \in A$ (where a is a name which does not appear earlier in the proof in any line one is allowed to use): in the context of this kind of proof it is appropriate to say "Let $a \in A$ be chosen arbitrarily" (and supply a line number so the assumption $a \in A$ can be used). One's goal is then to prove $P(a)$. Once the goal is achieved, one is entitled to assume $(\forall x \in A.P(x))$ and should not make further use of the lines that depend on the assumption $a \in A$. It is much more obvious in the restricted case that one gets a block of the proof that one should close off (because the block uses

a special assumption $a \in A$), and the restricted case is much more common in actual proofs.

These rules are called "universal generalization". The line reference would be to the block of statements from "Let $a[\in A]$ be chosen arbitrarily" to $P(a)$.

8.9.2 Using Universally Quantified Statements

If one is entitled to assume $(\forall x.P(x))$ and c is any name for an object, one is entitled to assume $P(c)$.

If one is entitled to assume $(\forall x \in A.P(x))$ and $c \in A$, one is entitled to assume $P(c)$.

These rules are called "universal instantiation". The reference is to the one or two previous lines used.

As rules of inference:

$$\frac{(\forall x.P(x))}{P(c)}$$

$$\frac{(\forall x \in A.P(x)) \qquad c \in A}{P(c)}$$

8.10 Existential Quantifier

This section presents rules for $(\exists x.P(x))$ ("for some x, $P(x)$", or equivalently "there exists an x such that $P(x)$") and for the restricted form $(\exists x \in A.P(x))$ ("for some x in the set A, $P(x)$" or "there exists x in A such that $P(x)$"). Notice that $(\exists x \in A.P(x))$ has just the rules one would expect from its logical equivalence to $(\exists x.x \in A \wedge P(x))$.

8.10.1 Proving Existentially Quantified Statements

To prove $(\exists x.P(x))$, find a name c such that $P(c)$ can be proved. It is your responsibility to figure out which c will work.

To prove $(\exists x \in A.P(x))$ find a name c such that $c \in A$ and $P(c)$ can be proved. It is your responsibility to figure out what c will work.

A way of phrasing either kind of proof is to express the goal as "Find c such that [$c \in A$ and] $P(c)$", where c is a new name which does not appear in the context: once a specific term t is identified as the correct value of c, one can then say "let $c = t$" to signal that one has found the right object. Of course this usage only makes sense if c has no prior meaning.

This rule is called "existential introduction". The reference is to the one or two lines used.

As rules of inference:

$$\frac{P(c)}{(\exists x.P(x))}$$

$$\frac{c \in A \qquad P(c)}{(\exists x \in A.P(x))}$$

8.10.2 Using Existentially Quantified Statements

Suppose that one is entitled to assume $(\exists x.P(x))$ and one is trying to prove a goal C. One is allowed to further assume $P(w)$ where w is a name which does not appear in any earlier line of the proof that one is allowed to use, and prove the goal C. Once the goal C is proved, one should no longer allow use of the block of variables in which the name w is declared (the reason for this is stylistic: one should be free to use the same variable w as a "witness" in a later part of the proof; this makes it safe to do so). If the statement one starts with is $(\exists x \in A.P(x))$ one may follow $P(w)$ with the additional assumption $w \in A$.

This rule is called "witness introduction" or "existential generalization". The reference is to the line $(\exists x[\in A].P(x))$ and the block of statements from $P(w)$ to C.

8.11 Proof Format

Given all these rules, what is a proof?

A proof is an argument which *can be* presented as a sequence of numbered statements. Each numbered statement is either justified by a list of earlier numbered statements and a rule of inference [for example, an appearance of B as line 17 might be justified by an appearance of A as line 3 and an appearance of $A \to B$ as line 12, using the rule of modus ponens] or is an assumption with an associated goal (the goal is not a numbered statement but a comment). Each assumption is followed in the sequence by an appearance of the associated goal as a numbered statement, which we will call the resolution of the assumption. The section of the proof consisting of an assumption, its resolution, and all the lines between them is closed off in the sense that no individual line in that section can be used to justify anything appearing in the proof after the resolution, nor can any assumption in that section be resolved by a line appearing in the proof after the resolution. In my preferred style of presenting these proofs, I will indent the section between an assumption and its resolution (and further indent smaller subsections within that section with their own assumptions and resolutions). The whole sequence of lines from the assumption to its resolution can be used to justify a later line (along with an appropriate rule of course): for example, the section of a proof between line 34: assume A: goal B and line 71: B could be used to justify line 113

$A \to B$ (lines 34-71, deduction); I do not usually do this (I usually write the statement to be proved by a subsection as a goal at the head of that section, and I do not usually use statements proved in such subsections later in the proof), but it is permitted.

I used to be in the habit of omitting the resolution of a goal if it was immediately preceded by an assumption-resolution section (or sections in the case of a biconditional) which could be used as its line justification: this seemed like a pointless repetition of the goal, which would already appear just above such a section. I would state the resolution line if it was going to be referred to in a later line justification. The idea was that the statement of a goal followed by a block of text that proves it is accepted as a proof of that statement; the only reason to repeat the statement with a line number is if it is going to be referenced using that line number. However, I have learned that students prefer closing lines.

Note the important italicized phrase "can be". A proof is generally presented in a mathematics book as a section of English text including math notation where needed. Some assumptions may be assumed to be understood by the reader. Some steps in reasoning may be omitted as "obvious". The logical structure will not be indicated explicitly by devices like line numbering and indentation; the author will rely more on the reader understanding what he or she is writing. This means that it is actually quite hard to specify exactly what will be accepted as a proof; the best teacher here is experience. A fully formalized proof can be specified (even to the level where a computer can recognize one and sometimes generate one on its own), but such proofs are generally rather long-winded.

8.12 Examples

These examples may include some general comments on how to write these proofs which you would not include if you were writing this proof yourself. I also included resolution lines (restatements of goals after they are proved) which I do not usually include.

Theorem: $((P \wedge Q) \rightarrow R) \leftrightarrow (P \rightarrow (Q \rightarrow R))$

Proof: The statement is a biconditional. The proof is in two parts.

Part 1: Assume (1) $(P \wedge Q) \rightarrow R)$

Goal: $(P \rightarrow (Q \rightarrow R)$

Now we use the strategy for proving an implication.

Assume (2) P

Goal: $Q \rightarrow R$

Assume (3) Q

Goal: R

Goal: $P \wedge Q$ (so that we can apply m.p. with line 1)

4 $P \wedge Q$ (from lines 2 and 3)

5 R rule of modus ponens with lines 1 and 4. This is the resolution of the goal at line 3.

6 $Q \rightarrow R$ lines 3-5. This is the resolution of the goal at line 2, which I used to omit.

7 $P \rightarrow (Q \rightarrow R)$ lines 2-6 This is the resolution of the goal at line 1, which I used to be in the habit of omitting.

Part 2: Assume (8): $P \to (Q \to R)$

Goal: $(P \wedge Q) \to R$

Assume (9): $P \wedge Q$

Goal: R

Goal: P (looking at line 1 and thinking of modus ponens)

10 P from line 9

11 $Q \to R$ mp lines 10 and 8.

Goal: Q (looking at line 4 and thinking of modus ponens)

12 Q from line 9

13 R lines 11 and 12, rule of modus ponens. This is the resolution of the goal at line 9.

14 $(P \wedge Q) \to R$ This is the resolution of the goal at line 8, which I used to omit.

15 $((P \wedge Q) \to R) \leftrightarrow (P \to (Q \to R))$ lines 1-14. I used to omit this as it just recapitulates the statement of the theorem already given. If I did omit it, I would also restart the numbering at 1 at the beginning of Part 2.

Theorem: $\neg(P \wedge Q) \leftrightarrow (\neg P \vee \neg Q)$

Proof: Part 1: Assume (1): $\neg(P \wedge Q)$

Goal: $\neg P \vee \neg Q$

We use the disjunction introduction strategy: assume the negation of one alternative and show that the other alternative follows.

Assume (2): $\neg\neg P$

Goal: $\neg Q$

Assume (3): Q

Goal: $\perp$ (a contradiction)

Goal: $P \wedge Q$ (in order to get a contradiction with line 1)

4 P double negation, line 3

5 $P \wedge Q$ (lines 3 and 4)

6 $\perp$ 1,5 contradiction . This resolves the goal at line 3.

7 $\neg Q$ lines 3-6 negation introduction. This resolves the goal at line 2.

8 $\neg P \vee \neg Q$ 2-7 disjunction introduction. This resolves the goal at line 1.

Part 2: Assume (9): $\neg P \vee \neg Q$

Goal: $\neg(P \wedge Q)$

Assume (10): $P \wedge Q$

Goal: $\bot$ (a contradiction)

We use the strategy of proof by cases on line 9.

Case 1 (9a): $\neg P$

Goal: $\bot$

11 : P from line 10

12 : $\bot$ 9a, 11 contradiction (this resolves the goal after 9a)

Case 2 (9b): $\neg Q$

Goal: $\bot$

13 Q from line 10

14 $\bot$ 9b, 13 contradiction (this resolves the goal after 9b)

15 $\bot$ 9, 9a-14 proof by cases.

16 $\neg(P \wedge Q)$ 9-15 negation introduction. This resolves the goal at line 9.

17 $\neg(P \wedge Q) \leftrightarrow (\neg P \vee \neg Q)$ 1-16, biconditional introduction.

Rule of Inference (Constructive Dilemma): We verify that

$$\frac{\begin{array}{r} P \vee Q \\ P \to R \\ Q \to S \end{array}}{R \vee S}$$

is a valid rule of inference.

If we are verifying a rule of inference we assume the hypotheses to be true then adopt the conclusion as our goal.

1 $P \vee Q$ premise

2 $P \to R$ premise

3 $Q \to S$ premise

Goal: $R \vee S$

We use proof by cases on line 1.

Case 1 (1a): P

Goal: $R \vee S$

4 R 1a, 2, modus ponens

5 $R \vee S$ addition, line 4. This resolves the goal at line 1a.

Case 2 (1b): Q

Goal: $R \vee S$

6 S 3,1b, modus ponens

7 $R \vee S$ addition, line 6. This resolves the goal at line 1b.

8 $R \vee S$ proof by cases, 1, 1a-7. And this is what we set out to prove.

Here is a quantifier example.

Theorem:

$$(\forall x : P[x]) \wedge (\forall y : P[y] \to Q[y]) \to (\forall z : Q[z])$$

Assume (1): $(\forall x : P[x]) \wedge (\forall y : P[y] \to Q[y])$

Goal: $(\forall z : Q[z])$

Let a be arbitrary.

(2): $a = a$ (optional)

Goal: $Q[a]$

(3): $(\forall x : P[x])$ simp 1

(4): $(\forall y : P[y] \to Q[y])$ simp 1

(5): $P[a]$ UI 3 $x := a$

(6): $P(a) \to Q(a)$ UI 4 $y := a$

(7): $Q[a]$ mp 5,6

(8): $(\forall z : Q[z])$ UG 2-7 [you may share my temptation to put a line number on "Let a be arbitrary" and start the UG block there; or, as shown here, use an optional line $a = a$ for this purpose]

(9): The theorem: deduction 1-8.

Chapter 9

Appendix: Description of the logic of Marcel

A *sequent* is a pair (p, g) where p is a finite sequence of formulas (the posits or premises in the sequent) and g is a finite sequence of formulas (the goals or conclusions in the sequent). The usual notation for (p, g) would look like this:

$$p_1, \ldots, p_n \vdash g_1, \ldots, g_m.$$

Notice that this looks slightly different from the treatment in chapter 5, where the paired objects are sets of formulas rather than finite sequences of formulas. Sequences have convenient structure for manipulation by a computer.

We note the syntax of propositional logic in Marcel: a proposition identifier is a string of lower case letters followed by a question mark. The symbols $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$ are replaced by

```
~, &, V, ->, == [note that the V is capitalized]
```

due to the limitations of the typewriter keyboard. Order of operations is as in the text, but conjunction and disjunction group by default to the left rather than to the right, for practical reasons having to do with what happens when large conjunctions or disjunctions are unpacked using the Marcel logical rules. It remains advantageous for implication to group to the right, and the biconditional groups to the right as well.

Notice also that in this notation posits are on the *left* and goals are on the *right*. This feature of terminology is preserved in Marcel though the Marcel display shows posits above and goals below:

```
1. p_1
.
.
.
n.   p_n

--------------------

1. g_1
.
.
.
m. g_m
```

Rules that act on posits or the list of posits are called left rules and rules that act on goals or the list of goals are called right rules.

We say that a sequent (p, g) is *valid* iff any assignment of meanings to non-logical symbols in formulas in the ranges of p and g which makes all the formulas in the range of p true makes at least one of the formulas in the range of g true.

The setting of Marcel which I use in teaching maintains the illusion that there are zero or more premises and exactly one goal in a sequent, which conforms to the way that arguments are usually presented.

The display looks like this:

```
1. p_1
.
.
.
n.   p_n

*2.  ~g_2
.
.
.
*m ~g_m

--------------------

1. g_1
```

If $m = 0$, so there are no goals, the format is

```
1. p_1
.
.
.
n.   p_n

--------------------

_|_
```

Observe that

$$p_1, \ldots, p_n \vdash g_1, \ldots, g_m$$

is valid if and only if

$$p_1, \ldots, p_n, \neg g_2, \ldots, \neg g_m \vdash g_1$$

is valid: to show that if all the premises p_i are true than some goal g_i is true is the same thing as to show that if all the premises p_i are true and all the goals $g_2, \ldots g_m$ are false, then the goal g_1 has to be true, and this appears to be the best presentation for students.

The order of posits and goals really does not matter, though they do have to be presented in some order (which is one reason that it is an advantage for Marcel to represent sequents using sequences).

The `gl` (get left) and `gr` (get right) commands allow the order of the posits and goals to be manipulated. If the sequent being viewed is

$$p_1, \ldots, p_n \vdash g_1, \ldots, g_m,$$

application of `gl(i)` will change it to

$$p_i, \ldots, p_n, p_1, \ldots, p_{i-1} \vdash g_1, \ldots, g_m,$$

and application of `gr(i)` will change it to

$$p_1, \ldots, p_n \vdash g_i, \ldots, g_m, g_1, \ldots, g_{i-1}$$

In the one-conclusion format `gr(i)` changes

$$p_1, \ldots, p_n, \neg g_2, \ldots, \neg g_m \vdash g_1$$

to

$$p_1, \ldots, p_n, \neg g_{i+1}, \ldots, \neg g_m, \neg g_1, \ldots, \neg g_{i-1} \vdash g_i.$$

If the sequent viewed is

$$p_1, \ldots, p_n \vdash g_1, \ldots, g_m,$$

and it happens that p_1 is the same formula as g_1 then the sequent is valid. If the `Done()` command is issued in this situation, Marcel records this sequent as proved and displays the next unproved sequent in the current proof (or declares the original theorem proved if no unproved sequents are left).

The previous command lets us finish things: we ought to report how we can get started: the command `s('p')` will set up the sequent $\vdash p$ to be proved valid, where p is any formula.

The very powerful `l()` (left) and `r()` (right) commands act on the first posit (left) or the first goal (right) applying the appropriate logical transformation based on the form of the leading posit or goal (as appropriate), producing either one or two new sequents to prove valid. Once these sequents are proved valid, the original sequent is recorded as proved valid.

conjunctions as posits:

$$P \wedge Q, p_2, \ldots, p_n \vdash g_1, \ldots, g_m$$

is valid iff

$$P, Q, p_2, \ldots, p_n \vdash g_1, \ldots, g_m$$

is valid. Thus, if the first posit in the sequent viewed is a conjunction, the effect of applying the `l()` command will be to break it apart into two posits.

conjunctions as goals:

$$p_1, \ldots, p_n \vdash P \wedge Q, g_2, \ldots, g_m$$

is valid iff both

$$p_1, \ldots, p_n \vdash P, g_2, \ldots, g_m$$

and

$$p_1, \ldots, p_n \vdash Q, g_2, \ldots, g_m$$

are valid. So if the first goal in the sequent (which in the one-conclusion format is the only goal below the line) is a conjunction, Marcel will first present the sequent with the conjunction replaced by its first part to prove, and then present the sequent with the conjunction replaced by the second part to prove: the strategy for proving $P \wedge Q$ under given assumptions is to first prove P with those assumptions, and then prove Q with those assumptions.

disjunctions as posits:

$$P \vee Q, p_2, \ldots, p_n \vdash g_1, \ldots, g_m$$

is valid iff

$$P, p_2, \ldots, p_n \vdash g_1, \ldots, g_m$$

is valid and

$$Q, p_2, \ldots, p_n \vdash g_1, \ldots, g_m$$

is valid. So if the first posit is a disjunction $P \vee Q$ and the `l()` command is applied, Marcel first presents the sequent with $P \vee Q$ replaced by P to be proved valid, then presents the sequent with $P \vee Q$ replaced by Q to be proved valid. This is exactly the strategy of proof by cases!

disjunctions as goals:

$$p_1, \ldots, p_n \vdash P \vee Q, g_2, \ldots g_m$$

is valid iff

$$p_1, \ldots, p_n \vdash P, Q, g_2, \ldots g_m$$

is valid. This is delightfully simple under the hood, but in the one-conclusion mode it looks slightly more complicated:

$$p_1, \ldots, p_n, \neg g_2, \ldots \neg g_m \vdash P \vee Q$$

is valid iff

$$p_1, \ldots, p_n, \neg Q, \neg g_2, \ldots \neg g_m \vdash P$$

is valid. Notice that this is one of the cases of alternative elimination. The other case is readily recovered by issuing the command `gr(2)` to make Q the chosen conclusion instead of P.

negations as posits: The sequent

$$\neg P, p_2, \ldots, p_n \vdash g_1 \ldots g_m$$

is valid iff

$$p_2, \ldots, p_n \vdash P, g_1 \ldots g_m$$

is valid. Like any rule which adds or removes a posit, this has less obvious effects in one-conclusion mode. The sequent

$$\neg P, p_2, \ldots, p_n, \neg g_2, \ldots \neg g_m \vdash g_1$$

is valid iff

$$p_2, \ldots, p_n, \neg g_1 \ldots \neg g_m \vdash P$$

is valid. The strategy presented is that one can prove a sequent with a posit $\neg P$ by instead proving from the other original hypotheses and the negation of the original first goal that P must be true.

negations as goals: The sequent

$$p_1, \ldots, p_n \vdash \neg P, g_2, \ldots, g_m$$

is valid iff the sequent

$$P, p_1, \ldots, p_n \vdash g_2, \ldots, g_m$$

is valid. In the one-conclusion mode, this has interesting effects:

$$p_1, \ldots, p_n, \neg g_2, \ldots, \neg g_m \vdash \neg P$$

is valid iff the sequent

$$P, p_1, \ldots, p_n, \neg g_3, \ldots, \neg g_m \vdash g_2$$

is valid. Notice that when the goal $\neg P$ is removed, the former second goal becomes first goal, and in one-conclusion mode this looks non-trivial. Of course, if there is no second goal, we will get the empty conclusion situation with the fake goal of $\bot$. This is the standard strategy of negation introduction: to prove that $\neg P$, assume P and deduce a contradiction. But the transformation in the one-conclusion mode makes it look like a contrapositive move: to prove $\neg P$ from the assumption $\neg g_2$, assume P and prove g_2. If there is no g_2, we get the usual strategy of negation introduction.

implications as goals: The sequent

$$p_1, \ldots, p_n \vdash P \to Q, g_1, \ldots, g_m$$

is valid iff

$$P, p_1, \ldots, p_n \vdash Q, g_1, \ldots, g_m$$

is valid. This gives Marcel an action exactly similar to our rule of deduction.

implications as posits: The sequent

$$P \to Q, p_2, \ldots, p_n \vdash g_1, \ldots, g_m$$

is valid iff both of the sequents

$$p_2, \ldots, p_n \vdash P, g_1, \ldots, g_m$$

and

$$Q, p_2, \ldots, p_n \vdash g_1, \ldots, g_m$$

are valid. This looks a little different in one-conclusion mode:

$$P \to Q, p_2, \ldots, p_n, \neg g_2, \ldots, \neg g_m \vdash g_1$$

is valid iff both of the sequents

$$p_2, \ldots, p_n, \neg g_1, \ldots, \neg g_m \vdash P$$

and

$$Q, p_2, \ldots, p_n, \neg g_2, \ldots, \neg g_m \vdash g_1$$

are valid. This rule is always the hardest one to follow. The problem is that we are used to using the posit $P \to Q$ together with P in the rule of modus ponens, but Marcel prefers to act on a single posit. The strategy implemented here is to attempt to show that P follows from the other hypotheses, then that the original conclusion follows from Q and the other hypotheses: if this works, it does show that the original conclusion follows from the original hypotheses: deduce P, apply modus ponens to get Q, then deduce the original goal. The additional option is provided of swapping P for the original goal in the first sequent: in this case the original conclusion follows without using the posit $P \to Q$ at all.

biconditionals as posits:

$$P \leftrightarrow Q, p_2, \ldots, p_n \vdash g_1, \ldots, g_m$$

is valid iff

$$P \to Q, Q \to P, p_2, \ldots, p_n \vdash g_1, \ldots, g_m$$

is valid. We choose to just break a biconditional apart as a conjunction is broken apart: the user can break apart whichever of the new implication posits they want to use.

biconditionals as goals:

$$p_1, \ldots, p_n \vdash P \leftrightarrow Q, g_1, \ldots, g_m$$

is valid iff

$$P, p_1, \ldots, p_n \vdash Q, g_1, \ldots, g_m$$

is valid and

$$Q, p_1, \ldots, p_n \vdash P, g_1, \ldots, g_m$$

is valid. This gives precisely the behavior we expect.

This essentially completes the rules for propositional logic (Marcel supports a couple of other less-used operators).

We note the syntax of quantification for Marcel. What we write $(\forall x.P[x])$, Marcel writes as `(A x : P[x])` (of course $P[x]$ is not Marcel notation: this stands in for the Marcel translation of $P[x]$). What we write $(\exists x.P[x])$, Marcel writes as `(E x : P[x])` (of course $P[x]$ is not Marcel notation: this stands in for the Marcel translation of $P[x]$). It is an interesting fact in the background that Marcel understands $(\forall x.P[x])$ as having the underlying form $\forall(\{x : P[x]\})$ and $(\exists x.P[x])$ as having the underlying form $\exists(\{x : P[x]\})$: for Marcel, set-builder notation is more basic. But this does not affect the way that quantifier rules are implemented.

We present the rules for handling quantified goals and posits.

universally quantified statements as goals: A sequent

$$p_1, \ldots, p_n \vdash (\forall x : A[x]), g_2, \ldots, g_m$$

is valid iff

$$p_1, \ldots, p_n \vdash A[a], g_2, \ldots, g_m$$

is valid, where a is an atomic constant not appearing anywhere in the original sequent. This precisely reproduces the rule of universal generalization. Marcel actually generates the term a by applying a fresh numerical index to the bound variable x (producing a variable `x_n` with n a fresh numerical index).

universally quantified statements as posits: A sequent

$$(\forall x : A[x]), p_2, \ldots, p_n \vdash g_1, \ldots, g_m$$

is valid iff

$$A[t], (\forall x : A[x]), p_2, \ldots, p_n \vdash g_1, \ldots, g_m,$$

where t is any term at all. What Marcel actually does is provide in place of t an "instantiable" `x$n`, n being a fresh index. Marcel then allows the instantiable `x$n` to be replaced at any later point, throughout

the entire current proof, with any term not containing any constant or instantiable with index $\geq n$. The advantage of "instantiables" is that it might become clear only later what the best value is to plug in for x, and further that the `l()` command can be used for universal posits: in old versions of Marcel, a different command was needed for universal posits and existential goals, which required the intended t as an argument. The fact that a copy of the universally quantified posit is preserved reflects the fact that more than one instance of the universally quantified posit might be wanted in a proof. It also preserves the precise equivalence of the validity of the original sequent and the validity of the new sequent.

existentially quantified statements as goals: A sequent

$$p_1, \ldots, p_n \vdash (\exists x : A[x]), g_2, \ldots, g_m$$

is valid iff

$$p_1, \ldots, p_n \vdash A[t], (\exists x : A[x]), g_2, \ldots, g_m$$

is valid, where t is any term. Application of the `r()` command introduces t as an instantiable `x$n`: this can later be assigned a value (throughout the current proof) as discussed above. The idea is that we then can handle existential posits with the parameter-free `r()` command like all other posits, and also that it may not become evident until later in the proof what the best witness is to choose. This looks a little different in the one-conclusion format:

$$p_1, \ldots, p_n, \neg g_2, \ldots, \neg g_m \vdash (\exists x : A[x])$$

is valid iff

$$p_1, \ldots, p_n, \neg(\exists x : A[x]), \neg g_2, \ldots, \neg g_m \vdash A[t]$$

is valid. To understand the preservation of the original goal as an alternative goal, consider that if we choose the wrong witness by mistake we can move the existential goal back into the first goal slot and instantiate it again. This can be important if different witnesses are to be chosen in different cases in an argument.

existentially quantified statements as posits: A sequent

$$(\exists x : A[x]), p_2, \ldots, p_n \vdash g_1, \ldots, g_n$$

is valid iff

$$A[a], p_2, \ldots, p_n \vdash g_1, \ldots, g_n$$

is valid, where a is a new atomic constant not appearing in the original sequent. Marcel implements a in the form `x_n`, where n is a fresh index. This implements the rule of witness introduction quite naturally.

The command that instantiates symbols `x$n` takes two forms: to replace `x$n` with t, issue the command `Inst('t','x$n')` or the alternative form `SU('x$n := t')`. The appearances of `t` of course are to be replaced by a possibly quite complex expression.

Index

addition (logical rule), 18
alternative elimination (disjunction introduction: logical rule), 18
and (conjunction), 15
and, misleading use of in connection with union, 16
and, other uses of the English word, 16
and/or, 17
atomic formula, 14

biconditional, 20
biconditional, proof strategies for, 20
biconditional, substitutions using the, 20
binary predicate, 14

chaining of relations, implicit conjunction via, 16
conjunction (and), 15
conjunction (logic rule), 16
contrapositive, proving the (logic rule), 19

deduction (logical rule), 19
disjunction introduction (alternative elimination: logical rule), 18
disjunction(and/or), 17

equality predicate, 15
equivalence, logical, 19

formula (contrasted with "sentence), 14
formula (grammatical sentence), 14
formula, atomic, 14

goal (statements), 13

if...then... (implication), 18
iff, 20
implication (if...then...), 18
implication, other English paraphrases of, 18
indirect proof (logical rule), 19
is, uses of English word, 14

letters, uses of, reviewed, 15
logical equivalence, 19

Marcel theorem prover, 12
membership predicate, 15
modus ponens (logic rule), 19
modus tollens (logic rule), 19

necessary, 18

or (inclusive, disjunction), 17
or, exclusive, 17, 20
ordered pair, used to eliminate higher arity predicates, 15

posit(ed statements), 13
predicate, binary, 14
predicate, unary, 14

predicates of arity higher than 2 not needed, 15
proof by cases (logical rule), 18
prove (a statement), 13

sentence (constrasted with "formula), 14
simplification (logic rule), 17
substitution using biconditionals, 20
sufficient, 18

term (noun phrase), 14
ternary predicates not needed, 15

unary predicate, 14

variables, 15
variables, sentence, capital letters used for, 15