Charles University in Prague

Faculty of Arts

MASTER'S THESIS

Bc. Martin Blahynka

Naive set theory with exclusive interpretation of quantifiers

Naivní teorie množin s výlučnou interpretací kvantifikátorů

Department of Logic

Supervisor of the master's thesis: Mgr. Vít Punčochář, Ph.D.

Study programme: Logic

Prague 2023

Acknowledgements: I am very grateful to my supervisor, Vít Punčochář, for all his help and support. I'd also like to thank to my two friends, R. J. and J. R., for their time and helpful comments.

I hereby declare that I carried out this thesis independently, that all the used literature and sources are properly cited, and that this thesis has not been used within any other study programme.

In Prague, May 10, 2023 Martin Blahynka

Abstract: Naive set theory can be formalised in first-order logic as a theory with one axiom (of extensionality) and one axiom schema (of unrestricted comprehension). It is widely known that this theory is inconsistent. What is less known is that a mere reinterpretation of the quantifiers in the schema of unrestricted comprehension blocks all the well-known paradoxes of naive set theory. This is the case when the quantifiers are interpreted exclusively, which is an idea that originates in Wittgenstein's Tractatus in the context of elimination of identity from logic. In the context of set theory, the idea was first used by Jaakko Hintikka thirty five years later. This thesis introduces and investigates the possibility of using exclusive interpretation of quantifiers to avoid paradoxes of naive set theory. The main criterion of success is consistency of the resulting theory. The main result of this thesis is the proof that the set theories, which use the idea of exclusive interpretation and which Hintikka left as possibly consistent, are inconsistent. The inconsistency is discussed in the context of Russell's vicious circle principle, which is found to be inadequate.

Abstrakt: Naivní teorii množin je možné formalizovat v logice prvního řádu jako teorii s jedním axiomem (extenzionality) a jedním axiomatickým schématem (neomezené komprehenze). Dobře známým faktem je, že taková teorie je sporná. Avšak méně známým faktem je to, že pouhá reinterpretace kvantifikátorů ve schématu neomezené komprehenze zablokuje všechny dobře známé paradoxy naivní teorie množin. Jde o exkluzivní interpretaci a tento nápad pochází z Wittgensteinova Traktátu, kde se objevuje v kontextu možnosti eliminace identity z logiky. V kontextu teorie množin jej poprvé použil až Jaakko Hintikka o třicet pět let později. Tato práce představuje a zkoumá možnost použití exkluzivní interpretace kvantifikátorů k zablokování paradoxů naivní teorie množin. Hledaná teorie by měla být především bezesporná. Hlavním výsledkem práce je důkaz toho, že teorie množin, které využívají tuto reinterpretaci kvantifikátorů a u kterých Hintikka nechal otázku bezespornosti otevřenou, jsou sporné. Spornost těchto teorií je diskutována v kontextu Russellova principu bludného kruhu, který je zhledán nedostatečným.

Keywords: Naive set theory, Exclusive interpretation, Quantifier, Inconsistency, Vicious circle

Klíčová slova: Naivní teorie množin, Exkluzivní interpretace, Kvantifikátor, Spornost, Bludný kruh

Contents

Introduction

Ludwig Wittgenstein in Tractatus famously asserted: “Identity of object I express by identity of sign, and not by using a sign for identity. Difference of objects I express by difference of signs.” [1, 5.53].

This idea, applied to quantifiers in first-order logic, leads to *exclusive interpretation of quantifiers*, where distinct bound variables must have distinct values. In classical logic, identity is needed to e.g. assert that there are exactly two objects satisfying some unary predicate P. To this end, one says that there is some a and there is some b such that a and b are distinct and $P(a)$ and $P(b)$ and that for every object c, if $P(c)$, then c is a or b. Formally, one can do so by the formula $\varphi \equiv \exists a(P(a) \wedge \exists b(a \neq b \wedge P(b) \wedge \neg\exists c(c \neq a \wedge c \neq b \wedge P(c))))$[1].

However, given exclusive interpretation, one does not need identity for this. One simply says that there is some a and there is some b such that $P(a)$ and $P(b)$ and there is no c such that $P(c)$. Formally: $\psi \equiv \exists a(P(a) \wedge \exists b(P(b) \wedge \neg\exists c P(c)))$. Assuming the quantifiers in φ are inclusive (i.e. standard) and quantifiers in ψ exclusive, the two formulae are equivalent in the sense that they are satisfied in the same structures of signatures with the unary predicate symbol 'P'. Exclusive interpretation is explained in detail in Section 2.1.

Jaakko Hintikka [2] added exclusive quantifiers to the standard first-logic with inclusive quantifiers and showed that there is a translation between formulae using inclusive quantifiers to formulae using exclusive quantifiers, and vice versa (the latter direction is exemplified by the formulae φ and ψ above). Furthermore, the formulae with exclusive quantifiers translated from formulae with inclusive quantifiers do not contain the sign for identity, which can therefore be eliminated by using exclusive quantifiers.

More recently, Kai Wehmeier [3] has addressed the topic and argued for the possibility and desirability of eliminating the identity sign from first-order logic. Consequently, Wittgenstein's claim that “The identity-sign is [...] not an essential constituent of conceptual notation.” [1, 5.533] has a much stronger support than many would believe.

This thesis, however, is focused mostly not on philosophical questions about identity but on the idea of using exclusive interpretation of quantifiers to “fix” naive set theory (henceforth just “Hintikka's idea” as it was introduced by Hintikka [2]).

Naive set theory can be formalised as a first-order theory in the standard set-theoretical language with only one axiom and one axiom schema. The axiom is the standard extenstionality axiom and the schema is unrestricted comprehension having the general form: $\exists S \forall x(x \in S \leftrightarrow \varphi)$[2]. Naive set theory is inconsistent, but as realised by Hintikka [2, pp. 239–241], the known paradoxes are avoided when the quantifiers in the comprehension are exclusive. In that case, Russell's property *is not a member of itself* gives rise (by comprehension) to a set S of all sets that are not members of themselves *except for S itself*, which may or may

[1] In this thesis I use '≡' as a symbol for identity at the meta level, in contrast to '=' – identity at the object level.

[2] The fact that S is capital has no special relevance, I just denote sets that are being defined by the comprehension by capitals.

not be a member of itself regardless of whether the property holds for it. This is because, due to exclusive interpretation, the quantifier $\forall x$ in the comprehension axiom excludes S from the range of its possible values. This axiom given by comprehension of $\varphi \equiv x \notin x$ with exclusive quantifiers is translated to standard logic as $\exists S \forall x(x \neq S \to (x \in S \leftrightarrow x \notin x))$.

From now on, I will use the term "exclusive set theory" for any set theory with extensionality and unrestricted comprehension in which exclusive interpretation is somehow used to try to avoid the paradoxes. There is a number of exclusive set theories and they differ in which parts of comprehension are exclusive and which are inclusive. In particular they can differ in the way they treat parameters in their comprehension schemas: which bound quantifiers (if any) exclude from their range of possible values the values of the parameters.

Although Hintikka [4] realised that one of these exclusive set theories is inconsistent, he left open the question of consistency of other exclusive set theories. Consistency of at least one of these theories would mean that a simple (and arguably quite natural) reinterpretation of quantifiers can fix naive set theory. In a sense, such a theory would be similar to Quine's New Foundations (introduced in Section 1.3) in that it would be based on the axiom of extensionality and the axiom schema of unrestricted comprehension of naive set theory and would thus not require additional axioms for power sets, unions, etc., as these would be entailed by the comprehension schema. It would also be a set theory with e.g. the universal set and Frege's numbers, which are "too big" to be sets in the most standard set theory – Zermelo-Fraenkel set theory (henceforth ZF).

The main result of this thesis is the negative resolution of Hintikka's open problem: the theories considered by him are all inconsistent. There are still some options not considered by Hintikka left as possibly consistent: in particular, the exclusive set theory without parameters and without identity introduced in Chapter 2. Overall it seems that allowing the use of parameters in the comprehension of an exclusive set theory leads to inconsistency, while not allowing them leads to a plausibly consistent but also plausibly unworkable theory.

Hintikka's idea was partly motivated by the vicious circle principle formulated by Russell first in a discussion with Poincaré (introduced in detail in Section 1.1.3). The existence of the paradoxes of naive set theory (and also some other related paradoxes) was blamed on the existence of vicious circles and the principle was supposed to prevent the vicious circles from appearing. However, among other problems with this principle, many exclusive set theories seem to implement the principle yet they are inconsistent. The exclusive set theories can, along with various paradoxes and other set theories, serve as a testing ground for alternative versions of the vicious circles principle. A valid vicious circle principle should be violated by the inconsistent set theories but not by the consistent ones. For an investigation about whether a theory violates a vicious circle principle and whether it contains paradoxes and whether these two things match, usefulness of the theory in mathematical practice can be ignored. This is why even set theories which are unworkable can be of some interest.

The thesis consists of four chapters. The first one motivates Hintikka's idea by introducing Russell's Vicious circle principle, Wittgenstein's problems with identity, and Quine's set theory New Foundations.

The second chapter introduces exclusive interpretation of quantifiers, Hintikka's idea, and various exclusive set theories in detail. It sets the stage for the next chapter.

The third chapter contains the new results. It shows that a particular exclusive set theory which seems rather weak is inconsistent. A consequence is drawn from this result to inconsistency of the family of set theories considered by Hintikka.

The fourth chapter discusses Russell's Vicious circle principle in light of inconsistency of exclusive set theories. It suggests there are two distinct problems with it, one of which explains why it does not guard the exclusive set theories from inconsistency.

The thesis also includes Appendix which contains succinct exposition of all the theories with some kind of unrestricted comprehension mentioned throughout this thesis.

1. Motivation of exclusive set theories

Hintikka's idea has in some sense two roots: one of them is Vicious circle principle formulated and defended by Bertrand Russell as the way to avoid paradoxes; the other is Ludwig Wittgenstein's idea of using exclusive interpretation to eliminate identity from logic. Hintikka's idea can be seen as using Wittgenstein's exclusive interpretation to implement Russell's Vicious circle principle in set theory. These are the topics of the first two sections of this chapter. The last section introduces W. V. Quine's set theory New Foundations which is in some ways similar to exclusive set theories.

1.1 History of the paradoxes and the vicious circle principle

The discovery of paradoxes in naive set theory was ensued by discussion of what exactly is to be blamed for their appearance. The debate has not exactly been settled; rather, several ideas have been developed into systems avoiding the paradoxes, but none of them with universal support. Some of the major solutions are: Russell's theory of types, Zermelo-Fraenkel set theory, and Quine's New Foundations. But before these alternative theories have been established, a discussion had taken place. Why does naive set theory contain paradoxes? What should be avoided in a better, non-naive set theory?

This section provides a brief overview of the history of Russell's paradox, Russell's first reactions to it, and the subsequent discussion of Russell and Poincaré in the first decade of the twentieth century. In this discussion, among other things Vicious circle principle (henceforth VCP) was formulated, which is relevant for Hintikka's idea.

1.1.1 Finding Russell's paradox

Russell discovered the paradox roughly in the following way (see [5]). He intuitively thought that there should be the universal set U (i.e. set of all sets) as it is the extension of the concept *anything*. If that is so, how does the power set of U – $\mathcal{P}(U)$ – look like? Clearly $\mathcal{P}(U) \subseteq U$ because U contains[1] all sets. Also $U \subseteq \mathcal{P}(U)$ because every set is also a subset of U. Thus $U = \mathcal{P}(U)$[2].

Now, $U = \mathcal{P}(U)$ cannot be reconciled with Cantor's theorem which says that $|S| < |\mathcal{P}(S)|$ for every set, thus also $|U| < |\mathcal{P}(U)|$. By Cantor's theorem there is no bijection from U to $\mathcal{P}(U)$, yet if $U = \mathcal{P}(U)$, there are bijections – the simplest one being identity (henceforth denoted as i). Thus if it is the case that $U = \mathcal{P}(U)$,

[1] I use "a contains b" as synonymous with "b is a member of a".

[2] These remarks can be understood either informally without a specific theory in mind, or in some set theory with the universal set, like New Foundations where indeed $U = \mathcal{P}(U)$. For now, it is best to follow Russell in thinking without any axiomatic theory, relying not on axioms but on intuition.

Cantor's theorem must not hold either in general, or at least for this particular case.

Let's now follow, as Russell did, the proof of Cantor's theorem that $\mathcal{P}(S) > S$ for the particular case of U, $\mathcal{P}(U)$, and i. The proof starts by assuming (for a contradiction) that there is a surjection f from S to $\mathcal{P}(S)$ (in our case i from U to $\mathcal{P}(U) = U$). The proof continues by considering the set $A = \{x \in S | x \notin f(x)\}$ (in our case $R = \{x | x \notin x\}$). The proof ends by deriving a contradiction: $A \in f(A)$ iff $A \notin f(A)$ (in our case $R \in R$ iff $R \notin R$).

Cantor's theorem is proved by invoking a diagonal set, which in case of U is Russell's set R. This is good news for someone who wants to have a set theory with the universal set because it leaves open the option of "fixing" naive set theory by somehow not allowing problematic formulae like "$x \notin x$" into comprehension and thereby losing Cantor's theorem (at least its general applicability to all the sets, including U). Indeed, there are such set theories with universal set, most prominently Quine's New Foundations.

As a sidenote, there is a dual of Russell's paradox – Russell's hypodox: $H = \{x | x \in x\}$, which is also problematic for naive set theory. Is $H \in H$, or $H \notin H$? Both can be true because if $H \in H$, then H does satisfy the membership criterion of H, and if $H \notin H$, it does not. This by itself does not make the theory inconsistent but it refutes the idea that every property has a corresponding set as its extension just as well as Russell's paradox does. A discussion of hypodoxes in general can be found in [6].

1.1.2 Russell's reaction to the paradox

Russell found the paradox in 1901 and before settling on the theory of types in around 1908 he explored various solutions to this paradox, to other set-theoretical paradoxes, and also to some related non-set-theoretical paradoxes.

The main directions of his thinking are outlined in [7] where he identifies three categories of a solution:

- The zigzag theory
- The theory of limitation of size
- The no class theory

Regarding the zigzag theory, the idea is that:

> [W]e start from the suggestion that propositional functions determine classes when they are fairly simple, and only fail to do so when they are complicated and recondite. [7, p. 38]

In my view, the terms "simple" and "complicated" are slightly misleading, as e.g. $x \in x$ seems rather simple. Simplicity is not what matters for a "propositional function"[3] to determine a class, at least not in any obvious sense of the term.

[3] Russell's conception of propositional functions is not very obvious and is addressed e.g. in [8, pp. 16–17]. In my commentary I follow a modern terminology, using instead the term "property" and often expressing properties in the standard notation of first-order logic.

In any case, the challenge for the zig-zag theory is to explain which propositional functions are valid (simple) and which are not. One solution to this challenge came only some thirty years later from Quine [9].

The second category of a solution that Russell identifies is limiting the size of sets.

> This theory is naturally suggested by the consideration of Burali-Forti's contradiction, as well as by certain general arguments tending to show that there is not (as in the zigzag theory) such a thing as the class of all entities. This theory naturally becomes particularized into the theory that a proper class must always be capable of being arranged in a well-ordered series ordinally similar to a segment of the series of ordinals in order of magnitude; this particular limitation being chosen so as to avoid Burali-Forti's contradiction. [7, p. 43]

Gödel [10, p. 453] observes that the zig-zag theory might be called "intensional" and the theory of limitation of size "extensional" since the former theory considers some properties invalid based on their intension (or meaning) while the latter on (the size of) their extension.

The last Russell's category is the no class theory where "...classes and relations are banished altogether." [7, p. 45]

Out of the three categories, the no class theory was Russell's favorite and was later developed by him in more detail. While in 1905 this is not yet clear as he sees major problems with it:

> The objections to the theory are (1) that it seems obvious to common sense that there are classes; (2) that a great part of Cantor's theory of the transfinite, including much that it is hard to doubt, is, so far as can be seen, invalid if there are no classes or relations; (3) that the working out of the theory is very complicated, and is on this account likely to contain errors, the removal of which would, for aught we know, render the theory inadequate to yield the results even of elementary arithmetic. [7, p. 45],

later his preference for this solution becomes clear:

> I have [...] discovered that it is possible to give an interpretation to all propositions which verbally employ classes, without assuming that there really are such things as classes at all [...] That it is meaningless [...] to regard a class as being or not being a member of itself, must be assumed for the avoidance of a more mathematical contradiction; but I cannot see that this could be meaningless if there were such things as classes. [11, p. 376]

A succinct explanation of the idea is in [12, p. 636, my translation]:

> The thesis of the no-class theory is that all significant propositions concerning classes can be regarded as propositions concerning all or some of their members, i.e., as terms which satisfy some propositional function $\varphi(x)$. I have found that the only propositions concerning

> classes that cannot be regarded in this way are propositions of the kind that give rise to contradictions. It is therefore natural to assume that classes are simply linguistic or symbolic abbreviations. For example, when we say, "Men are included in mortals," we seem to be making a judgment about the class of men collectively; but when we say, "All men are mortals," we are not necessarily assuming that there is a new entity, the class of men, in addition to all men individually.

Interestingly, as early as in 1905, Russell described three categories that are very much like what we have today: New Foundations is a theory in the vein of the zigzag theory, Zermelo-Fraenkel set theory limits the size of sets, and Russell's type theory and its descendants have the no class theory as their ancestor.

1.1.3 Vicious circle principle

This section presents the origins of VCP and also some later reactions to it – by Hintikka and Gödel.

Discussion of Russell and Poincaré

In the years 1905–1909, Poincaré and Russell discussed the problem in a series of papers [13, 7, 12, 14, 15]. In this section I introduce this discussion. Much more comprehensive introduction to this topic of early discussions of vicious circles can be found in [8]. A modern treatment of the topic of predicativity[4] in general can be found in Feferman's work, e.g. in [16], [17].

Poincaré [13] analysed the paradox of Jules Richard, which can be introduced as follows.

> Richard's paradox: Let S be the set of all real numbers that can be defined (by a finite number of words). S is countable because there is countably many finite definitions. Because S is countable, it can be ordered as a countable sequence. Then one can define a real number a with 0 as its integral part and its n-th decimal being 1 iff the n-th decimal of the n-th number in S is 8 or 9, and $p+1$ if the n-th decimal of the n-th number in S is $p < 8$.[5] a is different from every number in S, yet it was defined by a finite number of words, hence the paradox.

Then, in the paragraph named "La Vraie Solution", Poincaré [13, p. 307] gives his "true solution": we can only define S as the set of all numbers that can be defined *without introducing the notion of the set S itself.* Otherwise the definition of S is seen to contain a vicious circle.

He believed that many paradoxes, including the ones in set theory, exist because a vicious circle is somehow involved. The solution, then, is to avoid vicious circles in our definitions, i.e., to reject impredicative definitions. In fact, Poincaré identifies impredicative definitions with those that contain a vicious circle (see [13, p. 307]).

[4] There is clearly a strong connection between predicativity and vicious circles. As mentioned below, Poincaré identifies *predicative definitions* with those that do not contain a vicious circle.

[5] 8 is coupled with 9 just to avoid problems with $a = 0.9999... = 1$, $0.23999... = 0.24$, and the like.

While Poincaré is the first one to mention vicious circles in the context of the paradoxes, Russell [12] introduces VCP as a principle that is supposed to guard us against vicious circles. He writes:

> [T]he key to the paradoxes must lie in the idea of the vicious circle; I further recognise this to be true of Mr. Poincaré's objection to the idea of totality, that whatever in any way concerns *all* or *some* or *any* of the members of a class must not be a member of the class. In the language of Mr. Peano, the principle I hold may be stated as follows: 'Anything that contains an apparent variable must not be one of the possible values of that variable'. [12, p. 634, my translation]

Note that "apparent variable" is what we nowadays call "bound variable".

Russell was not very consistent with his formulations of VCP (see e.g. [8, pp. 3–4]). In this quotation alone, there seem to be two distinct principles, the former being stronger than the latter. Essentially, the latter narrows down "concerning members of a class *in any way*" of the former to "concerning members of a class by containing an apparent variable, the possible values of which are the members of the class". And as argued in Chapter 4, a broader sense of "concerning" is needed even in the context of set theories formalised in first-order logic.

A few pages later, Russell says:

> To avoid the fallacy of the vicious circle, we must admit [...] the principle: 'Everything that contains an apparent variable must be excluded from the possible values of this variable'. We will call this the principle of the vicious circle. [12, p. 640, my translation]

Thus it seems appropriate to use the term "Russell's VCP" for the following, even though there are other formulations by Russell himself.

> RUSSELL'S VCP: Everything that contains an apparent variable must be excluded from the possible values of this variable.

This formulation of Russell's VCP is uses the notion of "apparent variable" (i.e. bound variable). Such a formulation is inappropriate in the context of non-set-theoretical paradoxes like Richard's paradox or Liar's paradox which are usually introduced in a natural language, not in a formal one using variables and quantifiers. In such cases, some other Russell's formulations of the principle seem more appropriate, such as:

> RUSSELL'S VCP (INFORMAL): "Whatever involves all of a collection must not be one of the collection." [14, p. 225]

From now on I will use "RUSSELL'S VCP" and "RUSSELL'S VCP (INFORMAL)" for these two formulations, and "Russell's VCP" more generally, without a specific formulation in mind.

Although Poincaré and Russell agreed on the need to avoid vicious circles, they disagreed on why these vicious circles appear and thus how they should be avoided. This is because Poincaré blamed their appearance on misguided belief in actual infinity while Russell held that such a belief is innocuous.

Poincare's perspective was that:

> It is the belief in the existence of the actual infinity that has given rise to these non-predicative definitions. Let me explain: these definitions include the word *all*, as can be seen from the examples [e.g. of Richard's paradox]. The word *all* has a clear meaning when it concerns a finite number of objects; in order for it to still have a meaning when the objects are infinite in number, there would have to be an actual infinity. Otherwise, not *all* these objects can be conceived as posited prior to their definition, and then if the definition of a notion N depends on *all* the objects A, it may be tainted by a vicious circle, if among the objects A there are some that cannot be defined without involving the notion N itself. There is *no actual infinity*; the Cantorians have forgotten this, and have fallen into contradiction. [13, p. 316, my translation]

Poincaré's view of how the belief in actual infinity can introduce vicious circles is further elaborated by him on the first pages of [15]. Hintikka distinguishes VCP in Russell's sense from VCP in Poincaré's sense, the latter being stronger [2, pp. 244–245].

Russell explicitly denies that the belief in actual infinity would play such a role regarding the appearance of vicious circles:

> "[C]ontradictions have no essential relation to infinity. Of the *insolubilia* considered by the ancients, none introduces infinity; and it is singular that Mr. Poincaré cites Epimenides [i.e. Liar's paradox] as analogous to those which occur in the theory of the transfinite. A simplification of this paradox is constituted by the man who says: 'I lie'; if he lies, he tells the truth; but if he tells the truth, he lies. Has this man forgotten that there is no actual infinity?" [12, p. 633, my translation]

Another paradox that does not involve infinity mentioned by Russell (e.g. in [14]) is Berry's paradox. A version of it is:

> Berry's paradox: Define a as *the smallest natural number which does not have a definition of less than fifty syllables in English language.* The number a exists because there are only finitely many definitions with less than fifty syllables, so the class of all natural numbers without such a definition is non-empty and thus has the smallest number. However, this number has just been given a definition in less than fifty syllables – hence the paradox.

Before moving on to the next chapter, I briefly mention two reactions to this discussion of VCP – one by Gödel [10], the other by Hintkka [2]. The latter as a motivation of Chapter 2, the former as in some ways relevant for Chapter 4.

Gödel's analysis of VCP

Gödel claims that VCP, as formulated on several occasions by Russell, are in fact three different principles. Following Russell, Gödel formulates VCP as: "[N]o totality can contain members definable only in terms of this totality, or members involving or presupposing this totality.” [10, p. 454]

Then he notes that “corresponding to the phrases ‘definable only in terms of,’ ‘involving,’ and ‘presupposing,’ we have really three different principles, the second and third being much more plausible than the first.” [10, p. 455]

The main point of Gödel is that while VCP seems plausible for a constructivist, it does not seem plausible for a realist. Focusing mainly on the first form of VCP, which is the strongest, Gödel writes:

> [I]t seems that the vicious circle principle in its first form applies only if the entities involved are constructed by ourselves. In this case there must clearly exist a definition (namely the description of the construction) which does not refer to a totality to which the object defined belongs, because the construction of a thing can certainly not be based on a totality of things to which the thing to be constructed itself belongs. If, however, it is a question of objects that exist independently of our constructions, there is nothing in the least absurd in the existence of totalities containing members, which can be described (i.e., uniquely characterized) only by reference to this totality. [10, p. 456]

Regarding the other two forms of VCP, Gödel says that these do not seem to be valid, in general, for a realist either, because “one cannot say that an object described by reference to a totality ‘involves’ this totality, although the description itself does” and “nor would it contradict the third form, if ‘presuppose’ means ‘presuppose for the existence’ not ‘for the knowability.’ ” [10, p. 456] In the particular case of set theory where the objects are sets (or classes), Gödel is even willing to concede VCP in the second and third form but rejects it in the first form.

> As to classes in the sense of pluralities or totalities it would seem that they are [like concepts] not created but merely described by their definitions and that therefore the vicious circle principle in the first form does not apply. I even think there exist interpretations of the term ‘class’ (namely as a certain kind of structures), where it does not apply in the second form either. But for the development of all contemporary mathematics one may even assume that it does apply in the second form, which for classes as mere pluralities is, indeed, a very plausible assumption. [10, p. 459]

Hintikka on VCP

Hintikka’s idea (of interpreting quantifiers exclusively to avoid the paradoxes) is, in retrospect, almost forcibly suggested by some formulations of VCP by Russell, e.g. by Russell’s VCP and Russell’s VCP (informal) above.

As mentioned in Introduction, Hintikka first formulated the core of his idea in [2], noting that there are several possible implementations of this idea (i.e., several possible exclusive set theories) but focusing mainly on the most straightforward one. One year later he realised that that particular implementation of the idea was inconsistent and left open the question of consistency of the others, again with a particular emphasis on one of these. These two theories are introduced both in Chapter 2 and in Appendix and I call the first theory T_{HF} and the second theory T_{HS}.

Russell's aforementioned formulations of VCP prompted Hintikka to say:

> It may also be pointed out that the inconsistency of [T_{HF}] would have highly interesting consequences concerning certain principles used to guide the building up of various systems of mathematical logic. It may be argued that if [T_{HF}] gives rise to contradictions, then the celebrated vicious-circle principle is false in the sense that it does not, under an extremely natural interpretation of the principle, rule out all the paradoxes. This presupposes, obviously, that [T_{HF}] may be interpreted as a way of carrying out the vicious-circle principle. [2, p. 242]

It turns out that the system *is* inconsistent, hence Hintikka's view one year later is that "[The inconsistency of T_{HF}] means, in effect, that *the vicious circle principle is false* under a very natural interpretation of the principle." [4, p. 246].

Hintikka wrote this around fifty years after the time period in which Russell was dealing with these problems. Consequently, one can only speculate what Russell would think about this claim that the inconsistency of T_{HF} shows invalidity of Russell's VCP. I only have two remarks about Russell's position.

Firstly, it is not clear that what Hintikka calls "Russellian version of the vicious circle principle" or what I call "Russell's VCP" really is a good approximation of Russell's position (although calling it so is justified because it is based on Russell's own formulations of the principle). Recall that Russell formulated the principle on many occasions. While RUSSELL'S VCP forbids "containing an apparent value...", the more general formulations forbid "concerning in any way..." Arguably, T_{HF} (and other inconsistent exclusive set theories) violate some versions of Russell's principle but do not violate some other versions. Extending Hintikka's claim that Russell's VCP is invalid to the claim that Russell was wrong about what gives rise to vicious circles would at the very least require work. In short, perhaps he just was not very careful in some of the formulations.

Secondly, recall Poincaré's proposed solution to Richard's paradox. According to Poincaré, instead of a class S of those numbers which can be defined, we can only have a class S of those numbers which can be defined *without introducing S itself*. However, Russell did not agree that this approach avoids vicious circles:

> The method by which Mr. Poincaré tries to avoid the vicious circle consists in saying that when we assert 'All propositions are true or false', which is the law of excluded middle, we tacitly exclude the law of excluded middle itself. The difficulty is to legitimise this tacit exclusion without falling back into the vicious circle. [12, p. 644, my translation]

He says that we cannot define the law of excluded middle as "All propositions except the law of excluded middle are true or false." because the vicious circle in such a formulation is "flagrant".

> We must therefore find a way of formulating the law of excluded middle in such a way that it does not apply to itself, without saying, in formulating it, that it does not apply to itself. [12, p. 645, my translation]

Similarly, it is possible to see exclusive interpretation used in the exclusive set theories (including T_{HF}) as containing a vicious circle. When quantifiers in comprehension are exclusive, a definition of a set looks like: "Set S such that all sets, *except S itself*, are its members if and only if ..." Such a formulation might be said to contain a vicious circle, because S itself is mentioned in the definition of S.

However, this is not necessarily so. If one follows Wittgenstein in using logic with exclusive quantifiers, the object being defined or the proposition being asserted is excluded automatically, so to speak, without the need of mentioning it. "All propositions" in "All propositions are true or false" automatically excludes this very proposition. And as Hintikka [2, pp. 1–2] argues, we are used to such exclusive interpretations from natural language, e.g. when one says "Mazzini did more for the emancipation of his country than any living man of his time" – clearly this does not mean that Mazzini did more than Mazzini, but only that he did more than any *other* living man of his time.

Russell says that we must find a way to formulate the law of excluded middle *without saying that it does not apply to itself* and the logic with exclusive interpretation of quantifiers is arguably one such way.

In conclusion, I see Hintikka's idea as a natural implementation of Russell's VCP in set theory, although Russell's position on the matter is not entirely clear and may not be captured faithfully by Russell's VCP. Chapter 2 introduces Hintikka's idea of using exclusive interpretation to implement Russell's VCP and avoid paradoxes of naive set theory. There is more to be said about VCP in general but it is left to Chapter 4.

1.2 The problem with identity

Exclusive interpretation of quantifiers is used in this thesis to try to avoid paradoxes of naive set theory. However, it is more often discussed in the context of a different aim: eliminating identity from logic. It was also in this context that the idea of exclusive interpretation of quantifiers originated in Tractatus. Therefore, it seems appropriate to dedicate a section to the topic of identity, even though the rest of the thesis is concerned with avoiding set-theoretical paradoxes and by and large ignores the philosophical problems of identity (in fact, Hintikka's exclusive set theories use identity). This section introduces the origins of this idea in its first part and and its more recent development in the other part.

1.2.1 Wittgenstein and identity

Wittgenstein's problem with identity traces back at least to Frege:

> Now if we were to regard identity as a relation between that which the names 'a' and 'b' designate, it would seem that $a = b$ could not differ

> from $a = a$ (i.e., provided $a = b$ is true). A relation would thereby be expressed of a thing to itself, and indeed one in which each thing stands to itself but to no other thing. What is intended to be said by $a = b$ seems to be that the signs or names ‘a’ and ‘b’ designate the same thing[...] [18, p. 209]

The problem with identity is then raised by Russell eighteen years before Tractatus:

> The question whether identity is or is not a relation, and even whether there is such a concept at all, is not easy to answer. For, it may be said, identity cannot be a relation, since, where it is truly asserted, we have only one term, whereas two terms are required for a relation. And indeed identity, an objector may urge, cannot be anything at all: two terms plainly are not identical, and one term cannot be, for what is it identical with? [19, p. 65]

He nevertheless does not see how identity could be eliminated and concludes:

> Thus identity must be admitted, and the difficulty as to the two terms of a relation must be met by a sheer denial that two different terms are necessary. There must always be a referent and a relatum, but these need not be distinct; and where identity is affirmed, they are not so. [19, p. 65]

Wittgenstein discusses this problem in Tractatus in similar terms as Russell:

> Roughly speaking, to say of *two* things that they are identical is nonsense, and to say of *one* thing that it is identical with itself is to say nothing at all. [1, 5.5303]

It is important to differentiate the question of identity of objects (or “things” in Wittgenstein’s terminology or “terms” in Russell’s) from the question of co-reference of names. Consider the proposition: “The city called ‘Prague’ is the capital of the Czech Republic”. Although the question of how we should properly understand such statements has a rich history with diverse opinions, most philosophers would presumably understand such statements as asserting a relation between *names* and not *objects*, as Frege did in the quotation above. Thus one can say that “the city called ‘Prague’” and “the capital of the Czech Republic” are names and as names they are not identical. They are, however, related by the equivalence relation which could be called “co-reference”. This relation holds between two names iff they stand for the same object. Understanding identity as a relation between names was the view, among many others, of Carnap, Zermelo, Dedekind, and Frege in Begriffsschrift (although he later changed his position and saw identity as a relation between *senses*, see [20, p. 154]).

It is not in contention whether there is such an equivalence relation between names. What is in contention is whether there is an equivalence relation *identity* between objects which holds for two objects iff they are not two (distinct) objects but only one object.

Besides the problem with identity just described, Wittgenstein in Tractatus seems to have another problem with identity:

> [Wittgenstein] held that every proposition is a truth function of elementary propositions, where each elementary proposition indicates that objects are disposed to one another in a determinate way. [21, p. 141]

Consequently, he believed it is nonsensical to say, e.g., that "'There are objects', as one might say, 'There are books'. And it is just as impossible to say, 'There are 100 objects'[...]" [1, 4.1272] This is because propositions like "There are 100 objects" are not concerned with *how objects are disposed to one another* at all. However, in the standard logic with identity *it is possible* to say this.

For these reasons, Wittgenstein decides to eliminate identity from logic, and for this purpose he comes up with the idea of using exclusive interpretation:

> Identity of object I express by identity of sign, and not by using a sign for identity. Difference of objects I express by difference of signs. [1, 5.53]

This is meant to eradicate the need for identity in logical language. Wittgenstein in Tractatus gives some examples like:

> [...]'Only *one* x satisfies $f()$', will read '$(\exists x).fx$:$\sim(\exists x, y)$.fx.fy'. [1, 5.5321]

Note that ':' signifies conjunction and '$\sim$' negation.

Wittgenstein did not work out systematically translation between exclusive and inclusive quantifiers. This was done by Hintikka [2] and Wehmeier [3] who extends Hintikka's result also for languages with individual constants.

Importantly, note that exclusivity does not lead to the inability to assert that something is true for *any other* b *or for* a *itself* in a context where a is already in use. What is classically expressed by "There is some a such that for every b it is the case that $Q(a, b)$" can be expressed, given exclusive interpretation, by "There is some a such that for every b *distinct from* a it is the case that $Q(a, b)$, *and also* it is the case that $Q(a, a)$". When one's quantification excludes a, it is because a is already "in use" and thus things about a can be asserted separately.

1.2.2 A modern discussion of identity

A recent discussion of the possibility and desirability of eliminating identity from first-order logic is to be found in [3]. Regarding the desirability, Wehmeier essentially argues in the same vein as Wittgenstein does. Regarding the possibility of eliminating identity from first-order logic, Wehmeier distinguishes four categories of using identity and argues for the possibility separately for each of the categories. Note that such an elimination is considered successful only if the new logic without identity is not less expressive than the standard one.

Firstly, regarding atomic formulae with functional terms like $f(x, y) = t$, instead of asserting identity between the functional term $f(x, y)$ and the term t we can assert $eval(f, x, y, t)$ – i.e., that the function f with arguments x, y evaluates to t. Identity is not involved in this. Secondly and thirdly, there are atomic formulas like $x = y$ and $x = c$ where x, y are variables and c a constant.

Both cases are eliminated by exclusive interpretation. Lastly, identity between referents of two distinct constants can be understood as co-reference of these constants without the need for identity relation between objects. At the end of this argument, Wehmeier writes:

> This concludes the *Tractatus*-inspired argument for the dispensability of objectual identity with respect to first-order logic. Given the general translatability of [standard first-order logic with identity] into [first-order logic with exclusive interpretation, without identity and with co-reference relation for constants], it also follows that mathematics, at least to the extent that it's formalizable in [standard first-order logic with identity], can be carried out without invoking an objectual identity relation. [3, p. 765]

Wehmeier's paper has sparked a subsequent discussion about whether identity really is completely eliminated by Wehmeier's approach (see e.g. [22, 23]). However, focusing on it would be too much of a digression, the main focus of the thesis are exclusive set theories.

1.3 New Foundations

New Foundations (henceforth NF) is a set theory conceived by W. V. Quine [9]. I now briefly introduce this theory because some references to this theory are made throughout the thesis due to some features it shares with exclusive set theories. Among these features is the existence of "big" sets like the universal set.

However, no serious attempt at any comprehensive introduction to the topic is made[6]. The formal aspects of the theory are also included in Appendix.

1.3.1 Introduction of NF

NF is formalised in first-order logic in the standard set-theoretical language and is akin to naive set theory and exclusive set theories in that its only axioms are the axiom of extensionality and the instances of unrestricted comprehension schema. However, not all first-order formulae that are allowed in the comprehension of naive set theory are allowed in the comprehension of NF, but only *stratified formulae* are. We say that a formula φ is stratified iff there is an initial segment $S = \{0, 1, ..., k\}$ of natural numbers and a function σ from the set of all variables in φ to S such that:

(i) for every atomic formula $x = y$, we have $\sigma(x) = \sigma(y)$, and

(ii) for every atomic formula $x \in y$, we have $\sigma(y) = \sigma(x) + 1$.

Besides that, φ must also meet the standard criterion of having as free variables only $p_1, ..., p_n, x$ and not S if the comprehension schema is given as:

$$\forall p_1 \forall p_2 ... \forall p_n \exists S \forall x (x \in S \leftrightarrow \varphi(x, p_1, p_2, ..., p_n)).$$

[6] For the original paper see [9]. For a simple introduction see [24]. For a comprehensive and detailed exposition see [25].

Note that New Foundations is essentially a solution to the paradoxes in the vein of Russell's zig-zag theory. In particular, which comprehension axioms of naive set theory remain has nothing to do with how big the resulting set is. For example, the comprehension axiom:

$$\exists S \forall x(x \in S \leftrightarrow \top)$$[7]

entails the existence of the universal set U.

Note that numbers are formalised in NF in the classical way: the number n is the set of all sets with n members. In many ways, NF is faithful to the original conceptions of set theory.

Note also that Russell's formula $x \in x$ is not stratified and therefore cannot be used in the comprehension.

1.3.2 Consistency of NF

The big question about NF is its consistency (e.g. relative to ZF). Jensen [26] proved that a modification of NF called NFU, different essentially in that extensionality does not apply to the empty set, is consistent. Regarding NF, Randall Holmes [27] has a claimed proof of consistency, but it has not yet been confirmed by the community.

It is natural to ask whether it is necessary to require in the definition of stratified formula that for every subformula $x \in y$ it is the case that $\sigma(y) = \sigma(x) + 1$ instead of just $\sigma(y) > \sigma(x)$. If the former option is more in accordance with the theory of types, the latter is more in accordance with ZF, where a set from a hierarchical level $V_{\alpha+1}$ may contain sets from V_β for all levels $\beta \leq \alpha$. Because this question seems to be ignored by the literature introducing NF, let me mention why stratification cannot be defined in this way to potentially save the reader some time (but this is not very important and the rest of this section can be skipped without consequences).

I will say that a formula is "semi-stratified" if it satisfies the definition of stratified formula altered by requiring that $\sigma(y) > \sigma(x)$ instead of $\sigma(y) = \sigma(x)+1$. The current question is whether allowing semi-stratified instead of stratified formulae in the comprehension leads to inconsistency. And it does, for the following reason. Consider the semi-stratified formula $\psi(x) \equiv \exists y(y \in x \wedge \forall z(z \in x \leftrightarrow z \in y))$. This formula essentially says that there is some $y \in x$ such that $y = x$, because of extensionality. Not surprisingly, this leads to a paradox, as it should not be possible to say that a set is identical to its member.

For $\neg\psi$, the comprehension gives the set P:

$$\exists P \forall x(x \in S \leftrightarrow \neg\psi).$$

Now, does P satisfy $\neg\psi$? If so, then it should be a member of itself, but then P does not satisfy $\neg\psi$. On the other hand, if P does not satisfy $\neg\psi$, then there must be some $y \in P$ with the same members as P. In that case, also $y \in y$ and thus y does not satisfy ψ – contradiction with $y \in P$.

[7] Strictly speaking, I do not consider the symbol $\top$ to be a part of the language. Thus it can be viewed as an abbreviation for any stratified tautological formula, the simplest being $x = x$

1.3.3 Cantor's theorem

The existence of the universal set U raises the question of the validity of Cantor's theorem. As explained in Section 1.1.1, these two are irreconcilable. The answer to this question is that the set that must be used in the proof of Cantor's theorem in general does not exist – it cannot be defined by a stratified formula. Recall that for the particular case of U and identity, this set is Russell's set. Consequently, NF-theorists distinguish *cantorian* sets from *non-cantorian* sets where Cantor's theorem holds true only for the cantorian sets. Big sets like U are not cantorian. Not surprisingly, the question of the status of Cantor's theorem in NF is addressed already by Quine [28].

There has been considerable research done in New Foundations, and this theory can therefore serve in some sense as a model for exclusive set theories – as a set theory to which exclusive set theories might be compared.

Regarding Cantor's theorem, one should expect it to have a similar status in exclusive set theories. However, a difference is that while Russell's formula $\psi \equiv x \notin x$ is not allowed in the comprehension schema of NF, in exclusive set theories it *is allowed*. In exclusive set theories, the paradoxes of naive set theory are not solved by banning some formulae from the comprehension, but by reinterpreting them. Thus a natural place to look for a paradox in exclusive theories would be to look at Cantor's theorem. If it could be proved in general, this would lead to inconsistency with the fact that a universal set exists[8]. However, Cantor's proof does not go through in exclusive theories. In the case of U, Cantor's proof in naive set theory invokes Russell's set and derives a contradiction. Although in exclusive theories there is a Russell's set given by the Russell's formula ψ, it may or may not contain itself, thus no contradiction can be derived.

[8] Using any tautological formula in the comprehension of an exclusive set theory gives a set of all sets *possibly except itself*. Such a set may or may not contain itself, and there could even be a universal set which does and a universal set which does not, which is why I write "a universal set" instead of "the universal set". The discussion of Cantor's theorem in exclusive set theories in this section, before introducing exclusive set theories in detail, relies on the fact that the basic idea has already been explained in Introduction and that the reader can refer to the Appendix. Alternatively, this discussion can be skipped and revisited later, after having read Chapter 2.

2. Exclusive set theories

This chapter presents Hintikka's idea of using exclusive quantifiers to "fix" naive set theory. It introduces several exclusive set theories and explains how the well-known paradoxes of naive set theory are avoided by these theories.

2.1 Exclusive interpretation of quantifiers

It is time to introduce exclusive interpretation of quantifiers in more detail. The reader who still finds this level of detail insufficient is advised to look at HIntikka's original paper [2][1].

Hintikka distinguishes two kinds of exclusive quantifier – *strongly exclusive* and *weakly exclusive*. He adds both to classical first-order logic with identity alongside inclusive (i.e. standard) quantifiers. Syntactically, both types of exclusive quantifier obey the same rules as inclusive quantifiers, but their semantics is different.

The semantics of a strongly exclusive quantifier is different from inclusive quantifier in that its value range excludes, from all individuals in the universe, the values of all bound variables in whose scope this quantifier lies, and also the values of all free variables (one can consider their scope to be the whole formula in question). The semantics of a weakly exclusive quantifier is such that its value range excludes, from all individuals in the universe, the values of all variables which occur freely in the scope of this quantifier.

Before continuing, three notes seem to be in order.

Note that the question of whether a bound variable can share its value with an individual constant can be ignored because there are no constants in the set-theoretical language. I will also ignore the question of whether the values of two free variables in the formula can coincide, as Hintikka does[2].

Note that I (following Hintikka) introduced the quantifiers as added to the classical logic, thus leading to a logic with three different kinds of quantifier. One could, however, opt to have logic with only one of these. This is the approach of Wehmeier [3] – he considers logic with (weakly) exclusive quantifiers, without other kinds of quantifier and without identity.

Note that the possibility of eliminating identity by using exclusive interpretation of quantifiers mentioned in Section 1.2 apply equally to both kinds of exclusive quantifier.

> [E]verything expressible in terms of the inclusive quantifiers and identity may also be expressed by means of the weakly exclusive quantifiers *without using a special symbol for identity*. The same statement

[1] Interestingly, in this paper Hintikka cites Otakar Zich (the founder of Charles University's Logic Department) as the author of "[t]he most resolute attempt to carry out an exclusive interpretation of bound variables[...]" [2, p. 229]

[2] "There remains an ambiguity concerning the interpretation of free variables. Are we to allow the values of two different free variables to coincide? Different answers to this question give rise to a further distinction between different kinds of calculi. We shall not discuss the resulting complications, however; they do not give anything new in principle. One can build a predicate calculus by means of bound variables only." [2, p. 230 (footnote)]

> is easily seen to hold also for the strongly exclusive quantifiers. [2, p. 235]

From now on I will, by default, assume strongly exclusive interpretation of quantifiers (strongly exclusive quantifiers seem to be better suited for set theory than weakly exclusive ones, as should become apparent in Section 2.3.2.) Also, the term “exclusive quantifier” will sometimes be used synonymously with “strongly exclusive quantifier”. A reader used to inclusive quantifiers can translate all the formulae with strongly exclusive quantifiers to formulae with standard quantifiers according to the following rule[3]:

One should go through the formula from left to right and every time one meets an existential quantifier, one transforms the current formula of the form $A\exists x(Z)$ to $A\exists x(x \neq y_1 \wedge ... \wedge x \neq y_n \wedge Z)$, and for a universal quantifier, one transforms $A\forall x(Z)$ into $A\forall x((x \neq y_1 \wedge ... \wedge x \neq y_n) \rightarrow Z)$, where y_1, ..., y_n are all variables in whose scope x is.

For example, given strongly exclusive interpretation, the formula

$$\exists a P(a) \wedge \neg\exists a \exists b(P(a) \wedge P(b))$$

says that there is exactly one individual which has the property P (this is the example quoted from Tractatus in Section 1.2.1, only in modern notation). The formula

$$\forall x \forall y Q(x, y)$$

says that every two distinct individuals have the relation Q to each other but says nothing of whether $Q(x, x)$ for some x[4].

These formulae would be translated to

$$\exists a P(a) \wedge \neg\exists a \exists b(a \neq b \wedge P(a) \wedge P(b))$$

and

$$\forall x \forall y(x \neq y \rightarrow Q(x, y)).$$

One more example: the formula

$$\exists a \forall b(P(a, b) \vee \exists c(P(b, a) \vee P(c, b)))$$

would be translated to:

$$\exists a \forall b(b \neq a \rightarrow (P(a, b) \vee \exists c(c \neq a \wedge c \neq b \wedge (P(b, a) \vee P(c, b))))).$$

Note that in the case of *weakly* exclusive quantifiers, the translations in the three examples above (from *strongly* exclusive quantifiers) would be the same. A simple example of a formula with exclusive quantifiers which would be translated differently based on whether the quantifiers are weakly or strongly exclusive can be given as follows. Take the subformula from the example in Tractatus:

[3] It is the other direction – translating a formula with inclusive quantifiers to a formula with exclusive quantifiers – that requires more work.

[4] This can be said separately: $\forall x \forall y Q(x, y) \wedge Q(x, x)$.

$$\exists a \exists b (P(a) \wedge P(b))$$

and alter it to:

$$\exists a P(a) \wedge \exists b P(b).$$

While in the case of both inclusive and strongly exclusive quantifiers the two formulae are equivalent, it is not so in the case of weakly exclusive quantifiers, where a can share its value with b only in the latter formula.

2.2 Introduction of exclusive set theories

In this section I present two exclusive set theories. One comes from Hintikka and the other is the most natural one in the current setting – all quantifiers in it are strongly exclusive. The thesis includes Appendix clearly describing the various set theories from this thesis for a quick reference, although the thesis without this appendix should be self-contained.

2.2.1 Features common to all exclusive set theories

There are several features that are common to all exclusive theories considered in this thesis:

- It is a theory with one binary predicate symbol '∈' in first-order logic. (It may or may not include identity.)
- The theory has the axiom of extensionality[5]:

  $$\neg \exists x \exists y ((y \in x \leftrightarrow y \in y) \wedge (x \in x \leftrightarrow x \in y) \wedge (\forall z z \in x \leftrightarrow z \in y)).$$

- The theory has some kind of unrestricted comprehension of the form:

  $$\exists S \forall x (x \in S \leftrightarrow \varphi)$$

  as an axiom schema, using exclusivity to some extent: at the very least, the value of S is excluded from the ranges of values of $\forall x$ and of all bound variables in φ.

The theories in this thesis differ in the concrete form of the comprehension schema: which parts of it are exclusive, which inclusive, how parameters work, and whether identity is allowed.

Note that a consequence of the fact that the value of S is excluded from $\forall x$ is that no such axiom entails self-membership of S nor its negation.

Consider the following simple comprehension axiom:

$\exists S \forall x (x \in S \leftrightarrow \bot)$[6].

[5] In words, there are no two distinct sets with the same members. Recall that strongly exclusive interpretation is assumed unless stated otherwise.

[6] $\bot$ can be understood as an abbreviation for any contradictory formula in the language without identity, e.g. $x \in x \wedge x \notin x$. Similarly for $\bot$: e.g. $x \in x \vee x \notin x$

This axiom entails that there is a (at least one) set which has no members except possibly itself. One cannot expect (at least without some additional axiom) that for every comprehension axiom there is exactly one set satisfying it. In the case of this axiom, there might be S_0 which has no member (not even itself), S_1 which has one member – itself, and even some other sets like S_2 which also has one member – itself, but is distinct from S_1[7]. Consequently, I will say "a set given by the comprehension axiom" and not "the set..." meaning *any set satisfying this axiom.*

2.2.2 Hintikka's first exclusive theory T_{HF}

Hintikka proposed the general idea in [2] where he focused mainly on the theory that is introduced in this section. I call this theory T_{HF} (as Hintikka's First). One year later Hintikka published a follow-up paper where he shows that T_{HF} is inconsistent. This proof is sketched in Section 2.4.1.

T_{HF} is very similar to naive set theory, as its comprehension allows identity and parameters, and exclusivity is only minimal; its comprehension schema is:

> $\forall p_1 \forall p_2 ... \forall p_n \exists S \forall x (x \in S \leftrightarrow \varphi(x, p_1, p_2, ..., p_n))$, where φ does not contain S and all quantifiers are inclusive except that the value of S is excluded from $\forall x$ and also from all bound variables in φ.

This schema cannot be written using the three kinds of quantifier (inclusive, weakly exclusive, strongly exclusive) without identity. The value of S is excluded from the quantifier $\forall x$ and also from quantifiers in φ, thus these cannot be inclusive. But if they are exclusive, they in general exclude more than S, whether they be strongly or weakly exclusive.

This is not a problem for Hintikka and he formalises the theory in the standard first-order logic only with inclusive quantifiers by adding inequalities to the appropriate places in the comprehension schema. This is also the approach taken in Appendix. It *would* be a problem if there was no identity in the language: in this case, the description of the comprehension schema would be complicated. While, as mentioned in Section 1.2, there are reasons to try and eliminate identity, this is ignored in the context of set theory. In this context, the focus of both Hintikka and of this thesis is to try to use the idea of exclusivity to avoid set-theoretical paradoxes. Philosophical problems of identity are mostly put aside.

2.2.3 A very exclusive set theory without parameters T_{WP}

The theory introduced in this section will be called T_{WP} (as a theory Without Parameters). In contrast to T_{HF}, T_{WP} can be introduced very simply without identity. Its comprehension schema is:

[7] Such a situation is not unfamiliar to someone who has encountered a theory of non-well-founded sets, as e.g. in [29]. In such theories, the axiom of extensionality is usually replaced by an alternative axiom which identifies S_1 with S_2 based on the fact that they are structurally the same. However, in this thesis I use the standard axiom of extensionality, as Hintikka does. The main focus is the question of consistency and the existence of distinct but structurally same sets does not threaten consistency in any way.

$\exists S \forall x (x \in S \leftrightarrow \varphi(x))$, where the only free variable in φ is x, and '=' does not appear in φ, and all quantifiers are strongly exclusive.

Later, in Section 3.1, I will introduce the theory T_0 which is similar to T_{WP} but with parameters added to the comprehension schema. As shown in Theorem 1, T_0 has trivial models unless an additional axiom is added. For the same reason, T_{WP} also has the additional axiom:

Additional axiom: $\exists a \exists b \exists c \top$ (i.e., there are at least three sets).

In contrast to T_{HF}, T_{WP} has a very natural and simple description: only exclusive quantifiers are used and identity is not needed. There are also two things that make T_{WP} (arguably) quite safe: exclusivity is used in the full extent, and parameters are not allowed. I suspect T_{WP} is consistent but this is just a guess. For an introduction of T_{WP} (or other theories) in classical logic with inclusive quantifiers, see Appendix.

2.3 Avoiding paradoxes

With the two exclusive theories properly introduced, it is now time to address the question of how they avoid paradoxes.

2.3.1 Russell's paradox

It is easy to see how the theories avoid Russell's paradox. Russell's formula gives us a set R from the following comprehension axiom:

$$\exists R \forall x (x \in R \leftrightarrow x \notin x).$$

But as long as the value of R is excluded from $\forall x$, which is the case in both T_{HF} and T_{WP} (in fact, in all exclusive set theories), there is no paradox here. R may or may not contain itself and the comprehension axiom above is indifferent to it. Regarding all the other sets, however, R contains them iff they are not members of themselves.

Interestingly, this possibility of avoiding Russell's paradox was realised by Frege and the attempt for having a consistent theory by only excluding the value of S from $\forall x$ (and not from bound variables in φ) in the comprehension schema is nicknamed "Frege's way out" by Quine [30]. However, this does not avoid the paradox introduced in the next section.

For a brief discussion of how exclusive set theories avoid well-known paradoxes including Burali-Forti's, see [2, pp. 239–241].

2.3.2 Paradox of non-loopy sets

Consider another paradox that appears in naive set theory. Let L be the set given by:

$$\exists L \forall x (x \in L \leftrightarrow \neg \exists z (z \in x \wedge x \in z))$$

I shall call the property *being a member of one of my members* "loopiness". Looking at sets as graphs, a set is loopy iff there is a cycle of length 2 starting and ending in this set. L is supposed to be the set of all non-loopy sets.

But the existence of this set leads to a simple contradiction in naive set theory where we also have the set $\{L\}$. Does $\{L\} \in L$? If so, $\{L\}$ must (from the definition of L) not be loopy, but it is loopy via L. If $\{L\} \notin L$, $\{L\}$ must be (from the definition of L) loopy, but it only has one member, so it must be loopy via this member, so it must be that $\{L\} \in L$.

Exclusive interpretation, however, saves the theory from this paradox in the following way. The property that the members of L must satisfy is now not *being absolutely non-loopy* but rather *being non-loopy with the possible exception of a loop via L*. Therefore, even if $\{L\}$ exists (which is in general not guaranteed in exclusive set theories), L contains it as it simply satisfies the criterion. There is no paradox here.

There is a difference between L in T_{HF} and T_{WP} in whether the value of x is excluded from $\forall z$ in the comprehension axiom above: in T_{HF} it is not, in T_{WP} it is. Consequently, in T_{HF} such L also contains self-membered sets which are not loopy in the sense of containing a loop of length 2, while in T_{WP} these sets are not in L because the loop of length 1 "does not count". This, however, bears no relevance with respect to the paradox.

Note that a Russell's set can be understood as a set of non-1-loopy sets, L as a set of non-2-loopy sets, and it is natural to also consider sets of all non-n-loopy sets for $n > 2$. In the present context, however, this brings nothing new. Such sets are involved in paradoxes in naive set theory but not in exclusive set theories for the same reason as in the case of L.

While Russell's paradox is avoided because the value of the set S being defined is excluded from $\forall x$, the paradox of non-loopy sets is avoided because the value of the set S being defined is excluded from the quantifiers in φ. Recall that this is not the case in Frege's way out, which is why it does not avoid the paradox of non-loopy sets.

Now it can be seen why I took *strongly* exclusive interpretation of quantifiers as default, and not *weakly* exclusive interpretation. Given the former, T_{WP} is the most natural exclusive set theory and it avoids this paradox. If weakly exclusive interpretation was taken as default, the most natural exclusive set theory would be analogous to T_{WP} but with weakly exclusive interpretation of quantifiers. However, because L does not appear in the subformula $\neg\exists z(z \in x \land x \in z)$ in the comprehension axiom, L would not be excluded from $\exists z$. In general, the set S being defined must be excluded from all the subsequent quantifiers in the comprehension schema to avoid the well-known paradoxes, which is not the case when the exclusivity is "weak". In the context of set theory, *strongly* exclusive interpretation seems more appropriate.

2.4 Problems of T_{HF} and T_{WP}

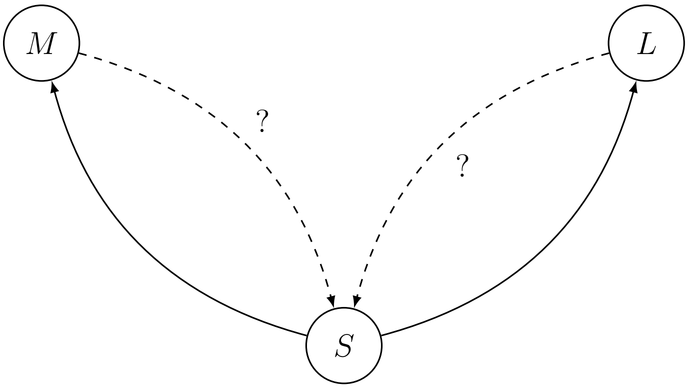
2.4.1 T_{HF} is inconsistent

Hintikka's theory T_{HF} manages to avoid the well-known paradoxes of naive set theory, but it is inconsistent. I now briefly sketch the contradiction found by

Hintikka [4].

Use comprehension (without parameters) to get L – a set of non-loopy sets, and M – a set of loopy sets. Note that L may contain loopy members if they are loopy only via L, as explained above. Then use L and M as parameters to define $S = \{L, M\}$ (or possibly $S = \{L, M, S\}$ but self-membership plays no role in the paradox). Now, if $S \in M$, S is loopy via something else than M – so $L \in S$ and $S \in L$. But the latter is true only if S is not loopy via something else than L, so $S \notin M$ – contradiction. If on the other hand $S \notin M$, then (from the definition of M) S is not loopy via L, thus $S \notin L$. But from the definition of L this implies that S is loopy via M, so $S \in M$ – again a contradiction. The situation is depicted in Figure 2.1.

Figure 2.1: Three sets L, S, M involved in the paradox. S defined as $\{L, M\}$ must contain both L and M. All possibilities regarding whether M contains S and whether L contains S lead to a contradiction.



This paradox is interesting in that the circularity in it is somewhat less immediate compared to Russell's paradox and the paradox of non-loopy sets. In those two paradoxes, there is one problematic set A (R in the case of Russell's paradox, L in the other paradox) and if A contains a certain set B (in Russell's paradox R itself, in the other paradox the singleton set $\{L\}$), B does not satisfy the defining property of A, and if A does not contain B, B *does* satisfy the property. The connection between $B \in A$ and B *satisfying the defining property of* A is immediate. In contrast, in this paradox, $S \in M$ does not immediately lead to S not satisfying the defining property of M. Instead, it immediately leads to S not satisfying the defining property *of* L. This is more properly addressed in Chapter 4 where it is argued that exclusive set theories in general avoid the "more immediate" circularity, but are still circular.

2.4.2 T_{WP} seems unworkable

While the problem of T_{HF} is inconsistency, the problem of the safer theory T_{WP} is that it is very non-classical in some ways. The problems become apparent already in very simple cases. Consider the following two comprehension axioms:

C_0: $\exists 0 \forall x (x \in 0 \leftrightarrow \bot)$,

C_1: $\exists 1 \forall x(x \in 1 \leftrightarrow \neg \exists z z \in x)$.

In a classical set theory, C_0 would entail the existence of the empty set and C_1 would entail the existence of $\{\emptyset\}$ – von Neumann ordinal 1. However, this is not necessarily so in T_{WP}. Regarding C_0, a set 0 satisfying it might have a member – itself. This ambiguity regarding self-membership is present in all exclusive set theories and arguably does not lead to problems. What is problematic is that the definition of C_1 does not necessarily entail the existence of a new set – i.e., a set distinct from 0. Suppose there is only one 0 satisfying C_0 – e.g. the classical empty set. Then 0 also satisfies C_1. This is because, after substituting the empty set 0 for 1 in C_1, the axiom is satisfied. It says that every *other* set is a member of 0 iff it is empty and it might be that there is no such set. In other words, it might be that the only set satisfying the property is 0 itself, but 0 is excluded from the possible values of $\forall x$ in C_1 if it is picked as the value of $\exists 1$.

Another problem is that a Frege's number F_n (i.e. the set of sets with n members) defined as

$$\exists F_n \forall x(x \in F_n \leftrightarrow (\exists a_0 \exists a_1 ... \exists a_{n-1}(a_0 \in x \wedge a_1 \in x \wedge ... \wedge a_{n-1} \in x) \wedge \neg \exists a_0 \exists a_1 ... \exists a_n(a_0 \in x \wedge a_1 \in x \wedge ... \wedge a_n \in x))$$

may in fact have members with different amount of members than n. For example, there might be a member m of F_n with $n+2$ members: m and F_n as extra members which "do not count" due to exclusivity. This might lead to problems if Frege's numbers are used to formalise numbers as was done historically by Frege and Russell, and as is usual in some set theories with big sets like New Foundations.

Regarding the classical operations union and intersection, they are also problematic. Given sets a and b by comprehensions on formulae φ_a and φ_b, the comprehension axiom given by $\varphi_a \wedge \varphi_b$ gives a set which is not necessarily the intersection of a and b.

Consequently, it is not clear how mathematics could be done in T_{WP}. One idea to solve some of these problems would be leaving comprehension the way it is and adding another axiom to the theory. In particular, adding axiom of pairing would seem to help because it would give us pairs and singletons. Then ordered pairs could be defined in a standard way. However, adding the axiom of pairing would lead to the very same contradiction that is present in T_{HF} and presented in the previous section. Another idea to solve some of the problems would be adding parameters. Note however that the values of the parameters must be excluded at least from $\forall x$ to avoid the paradox of T_{HF}. Consequently, parameters will not give us singletons and pairs. However, adding them might help with some other problems – like that of union and intersection. Adding parameters to T_{WP} leads to the theory T_0 introduced in Chapter 3.

2.5 Hintikka's open problem

T_{HF} is inconsistent and an obvious way to improve it is to exclude the values of parameters from the range of at least some quantifiers in the comprehension schema. Hintikka, after realising that T_{HF} was inconsistent, considered one such a theory which I call T_{HS}:

> The ordinary 'Russellian' version of the vicious circle principle prohibits the definition of an object or set x by means of totalities to which x itself belongs, *i.e.*, by means of bound variables one of the possible values of which is x itself. In $[T_{HS}]$, the range of the bound variables that may occur in the definition of x must not contain x nor one of the values of the free variables occurring therein. This means that the stronger principle formalized by $[T_{HS}]$ prohibits the definition of x by means of totalities that contain x *or one of the constant individuals with the reference to which x is defined.* [4, p. 249]

Note that x in the quotation corresponds to S in the general form of the comprehension schema used in this thesis. The theory T_{HS} is just like T_{HF} with only one difference: the values of parameters are also excluded from $\forall x$.

It seems surprising that only $\forall x$ excludes the values of the parameters. Hintikka defines T_{HS} this way even though the quotation above suggests that also the quantifiers used in φ should exclude them. Presumably, he leaves the quantifiers in φ as they are in T_{HF} because he realises that it does not matter whether they exclude the parameters or not – recall that exclusive quantifiers can be translated to inclusive ones, and vice versa, as long as identity is in the language (which it is in theories considered by Hintikka). Consequently, T_{HS} is equivalent to the theory T' which is like T_{HS} except also excluding the values of the parameters from quantifiers in φ in the comprehension.

To illustrate this with an example, suppose that

$$\varphi(x, p) \equiv \exists z(z \in x \wedge p \in z)$$

is used in the comprehension of T'. If φ was used in the comprehension of T_{HS}, the resulting axiom would have a different meaning because $\exists z$ includes the value of p among its possible values in T_{HS} but not in T'. But one can translate this formula φ to

$$\psi \equiv \exists z(z \neq p \wedge z \in x \wedge p \in z).$$

The comprehension axiom given by ψ in T_{HS} is equivalent to the comprehension axiom given by φ in T'.

In the other direction, if one starts with this formula $\varphi(x, p) \equiv \exists z(z \in x \wedge p \in z)$ in T_{HS}, one can translate it to

$$\psi' \equiv \exists z(z \in x \wedge p \in z) \vee (p \in x \wedge p \in p).$$

The possibility of such translations in general shows that T_{HS} is equivalent to T'. A similar argument cannot be made, of course, regarding exclusivity between $\exists S$ and $\forall x$, or between $\exists S$ and quantifiers in φ, because S cannot even appear in φ. Nor can such an argument show that T_{HF} is equivalent to T_{HS} because $\forall x$ does not appear in φ thus exclusivity between a parameter p and $\forall x$ cannot be "addressed" in φ.

Note that excluding the value of parameters from $\forall x$ has as a consequence that e.g. singletons do not necessarily exist. Hintikka comments:

> Among other things, $[T_{HS}]$ does not afford the usual definitions of a unit set, a couple, a triple, etc. [...] This, it seems, is part of the price

> we have to pay for the absence of the usual restrictions in terms of stratification or limitation in size. [4, p. 249]

Hintikka left open the question of whether T_{HS} is consistent. And as far as I know, no one has addressed this question. As proved in the next chapter, it is inconsistent.

3. New results

In this chapter I introduce the theory T_0 which seems rather safe. Compared to T_{WP}, parameters are added in the safest possible way. Compared to T_{HF}, the values of parameters are not included in the ranges of possible values of $\forall x$ and of quantifiers in φ.

After introducing T_0 at the beginning, its inconsistency is proved. Then it is argued that this inconsistency entails inconsistency of all exclusive set theories considered by Hintikka and left by him as possibly consistent.

3.1 A very exclusive set theory with parameters T_0

The theory T_0 is formulated to be, in some sense, as weak as possible: its comprehension schema does not allow identity and it is very exclusive. This complicates the proof of its inconsistency. However, it is so on purpose: once it is proved that even T_0 is inconsistent, inconsistency of many other exclusive set theories follows.

T_0 can be viewed as T_{WP} to which parameters are added in the safest possible way. The values of the parameters are excluded from all subsequent quantifiers in the comprehension schema except for $\exists S$. This exception is also made to make the theory as weak as possible. Note that if the values of the parameters were excluded from $\exists S$, there would have to be S distinct from all the values of the parameters for any number and choice of parameters. One consequence of this would be that there must be infinitely many sets, similarly to a claim from Tractatus:

> [...] What the axiom of infinity is intended to say would express itself in language through the existence of infinitely many names with different meanings. [1, 5.535]

3.1.1 Introduction of T_0

In the theory T_0 everything is exclusive except that the set S defined by the comprehension may be identical to the value of a parameter. The comprehension schema of T_0 is:

> $\forall p_1 ... \forall p_n \exists^i S \forall x (x \in S \leftrightarrow \varphi(x, p_1, ..., p_n))$, where φ does not contain 'S', nor '$=$', and all quantifiers are exclusive except for $\exists^i S$ which is inclusive.

Thus the only exception to exclusivity is that S can share a value with a parameter.

The theory T_0 has one extra axiom (added to extensionality and comprehension schema):

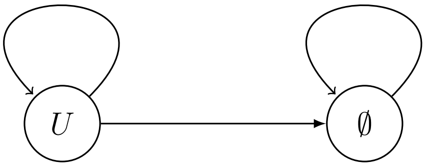
> *Additional axiom*: There are at least 3 sets.

This is because the theory T'_0, which is just like T_0 except without this additional axiom, actually has models with 1 or 2 sets in the universe.

T'_0 has a model with one set in the universe as there is nothing to force the existence of several distinct sets and all comprehension axioms are trivially satisfied if there is only one set in the universe.

Also, T'_0 has a model with two sets.

Figure 3.1: A model of T'_0 with two sets in the universe. The arrows depict the relation *contains* – the inverse of *membership*.



It can be proved that this is a model of T'_0[1].

Theorem 1. *The structure from Figure 3.1 is a model of T'_0.*

Proof. Extensionality is satisfied because the two sets do not have the same members: $U \in U$ but $U \notin \emptyset$.

Regarding all the instances of the comprehension schema, consider first the instances without parameters. All of them are of the form $\exists S \forall x(x \in S \leftrightarrow \varphi(x))$. Due to the exclusivity of the quantifier $\forall x$ in the axioms and the fact that there are only two sets, there are only two possible evaluations: either $S := U$ and $x := \emptyset$, or vice versa. I shall prove by induction on $\varphi(x)$ that for every $\varphi(x)$ either $\varphi(x)$ is true in both cases, or $\varphi(x)$ is false in both cases.

With this proved, it follows that all the instances of the comprehension without parameters are satisfied. If φ is true in both cases, $S := U$ will do, and if φ is false in both cases, $S := \emptyset$ will do.

If φ has no quantifiers, the only atomic formula that φ can be is $x \in x$ (as S cannot be mentioned in φ and we do not have idenity in our language[2]). In both cases this φ is true (because both U and $\emptyset$ are members of themselves).

If $\varphi \equiv \forall z(...)$ then it is trivially true in both cases due to the exclusivity because there is no z other than S and x.

If $\varphi \equiv \exists z(...)$ then it is trivially false in both cases because there is no z other than S and x.

If $\varphi \equiv \neg\psi$ then it is false in both cases if ψ is true in both cases and vice versa.

If $\varphi \equiv \psi_1 \vee \psi_2$ then it is false in both cases iff ψ_1 and ψ_2 are false in both cases, otherwise it is true in both cases.

[1] It can be proved analogously that a similar structure with two objects – identical to the one depicted except without the two loops – is also a model.

[2] If we did, the proof would work just in the same way. The atomic formula $x = x$ is a tautology and equivalent to e.g. $x \in x \vee x \notin x$.

To finish the proof we need to address the parameters. Firstly note that if we use two parameters then the axiom is trivially true because whichever value we pick for S (U or $\emptyset$, recall that the quantifier $\exists S$ is inclusive with regard to the parameters), $\forall x(...)$ is trivially true because there is no other x distinct from both values of the parameters. If there are three or more parameters, then already $\forall p_1 \forall p_2 \forall p_3(...)$ is trivially true because there is no third value for p_3.

But similarly, if there is only one parameter, we can pick the other value for S and the formula is again trivially true because there is no third value for x. □

I shall now turn to the proof of inconsistency of T_0. To appreciate some of the complications in trying to prove this, consider a set of non-loopy sets L defined by:

$$\exists L \forall x(x \in L \leftrightarrow \neg \exists z(z \in x \wedge x \in z)).$$

Could it be identical to $\emptyset$? Could it be identical to U (i.e. to the universal set given by a tautology in comprehension)? Neither of the two questions is trivial. If $L = \emptyset$, then it is no problem that $\emptyset \notin L$ even though $\emptyset$ is non-loopy, because this criterion is only used for sets other than L. So to show that this cannot be the case we would need another set that we know is non-loopy. But is there any such set? Parameters do not seem to help (even if we could use identity, which we cannot in T_0). E.g. a set S given by

$$\exists S \forall x(x \in S \leftrightarrow x = \emptyset)$$

could still be just $\emptyset$ because it is trivially true for $\emptyset$ that every other set belongs to it iff it is identical to $\emptyset$.

To show that $L \neq U$ it does not suffice to find some set S that contains U because if $L = U$ then it may still contain S as long as S does not contain another loop than the one via U. We would need to find two loopy sets distinct from U.

3.1.2 T_0 is inconsistent

The core idea of the inconsistency proof is the following. Start with a set of non-loopy sets L and use it as a parameter to define the sets A and B that are distinct from L and from each other but contain almost the same members (except for a few "irrelevant" ones, which makes them all distinct by extensionality). Then show that $A \in L$, $B \in L$: because if e.g. $A \notin L$, A would have to be loopy via some set C, in which case C would be loopy via A. But as long as C is not one of the "irrelevant" sets, $C \in A$ implies $C \in L$ which is a contradiction – L would contain a loopy set.

Then, as long as both A and B are not among the few "irrelevant" sets, $A \in L$ implies $A \in B$, and $B \in L$ implies $B \in A$. But then A is loopy via B and B is loopy via A and they are members of L, which is a contradiction – L only contains non-loopy sets.

Recall that we have a first-order theory with one binary symbol '$\in$' without identity and with exclusive interpretation of quantifiers. We have three axioms/schemas:

- $\neg\exists x \exists y ((y \in x \leftrightarrow y \in y) \wedge (x \in x \leftrightarrow x \in y) \wedge (\forall z (z \in x \leftrightarrow z \in y)))$ (extensionality)

- Comprehension schema introduced in the previous section

- $\exists a \exists b \exists c \top$ (i.e. there are at least three sets)

Let's start with the proof of inconsistency. Suppose for a contradiction that T_0 is consistent, thus has a model. By investigating how such a model must look like (which sets necessarily exist in it by comprehension and what the relations among them are), a contradiction is derived in the end.

To begin with, I list the sets that will be needed for the proof. L, A and B are needed for the core of the proof as sketched above, others can be thought of as auxiliary.

$\exists \emptyset \forall x (x \in \emptyset \leftrightarrow \bot)$
$\exists U \forall x (x \in U \leftrightarrow \top)$
$\exists L \forall x (x \in L \leftrightarrow \neg \exists z (z \in x \wedge x \in z \wedge \exists y y \notin z))$[3]
$\exists M \forall x (x \in M \leftrightarrow \exists z (z \in x \wedge x \in z))$
$\exists V \forall x (x \in V \leftrightarrow \forall y y \in x)$
$\exists W \forall x (x \in W \leftrightarrow (\forall y y \in x \vee \forall y y \notin x))$
$\exists D_1 \forall x (x \in D_1 \leftrightarrow \exists z z \in x)$
$\exists E \forall x (x \in E \leftrightarrow (\forall y y \in x \vee \forall y (y \in x \rightarrow \exists z z \in y)))$
$\exists A \forall x (x \in A \leftrightarrow (x \in L \vee \forall y y \in x))$ (L is a parameter)
$\exists B \forall x (x \in B \leftrightarrow (x \in L \wedge \exists y y \in x))$ (L is a parameter)

Note that the distinctness of the definitions does not imply distinctness of the sets – this will have to be proved when needed. And this is often needed: when one wants to prove that $A \in B$, one must first prove that $A \neq B$, because otherwise $A \in B$ is equivalent to $A \in A$ and no comprehension axiom entails self-membership nor its negation.

Also note that for every comprehension axiom above, there might be several distinct sets satisfying it. For instance, there might be U_1 – the set of *all sets* and U_2 – the set of all sets *except itself*. The comprehension axioms entail the existence of *at least one* such set. In such a case, the set U in the proof is any such set.

The first part of the proof (before parameters are used, to introduce A and B) can be understood as investigating how a model of T_{WP} must look like (it seems plausible to me T_{WP} is consistent). The other part, then, introduces A and B by using parameters in comprehension and derives a contradiction.

The sets and the membership relations among them that are proved in the following lemmas are depicted in Figure 3.2 for a better orientation in the proof.

Lemma 2. *$\emptyset \neq U$, $\emptyset \in U$, $U \notin \emptyset$.*

[3] Note that this definition of L is different from the definition of L earlier in this paper, due to the third conjunct.

Proof. The inequality follows from the fact that there are at least two sets as follows. Suppose that $U = \emptyset$ and let e be some other set. We have the contradiction: $e \in \emptyset$ (from the definition of U and the fact that $U = \emptyset$ (and the fact that $e \neq U$)) but also $e \notin \emptyset$ (from the definition of $\emptyset$ (and the fact that $e \neq \emptyset$)). $\emptyset \in U$, $U \notin \emptyset$ then simply follow from the definitions of the two sets. □

Lemma 3. *The sets $\emptyset$, U, V, and W are pairwise distinct and $\emptyset \notin V$, $\emptyset \in W$, $U \in V$, $U \in W$, and $V \notin W$.*

Proof. Suppose $V = U$ and let e be any set other than U and $\emptyset$. We have $\emptyset \in V$ (because $V = U$ and U contains every other set) but also $\emptyset \notin V$ (from the definition of V because $e \notin \emptyset$). Thus $V \neq U$. Then clearly $V \in U$ and $U \in V$ from their definitions (note that V might contain some other sets too – e.g. the set of all sets except V if such a set exists). $V \neq \emptyset$ because $U \in V$. $V \notin \emptyset$ is clear, and $\emptyset \notin V$ is from the fact that $U \notin \emptyset$ together with the definition of V.

Now consider W. $W = V$ would be a contradiction: $\emptyset \notin V$ but clearly $\emptyset \in W$. $W = \emptyset$ would be a contradiction: $U \in W$ but $U \notin \emptyset$.[4]

$W = U$ would also be a contradiction, although somewhat non-trivially. Suppose $W = U$. In that case, because $V \in W$ and $\emptyset \notin V$, V must contain no set, except possibly W and V itself, to satisfy the criterion of W (the second disjunct in the definition of W). Now consider M. $M \neq \emptyset$ (otherwise there would be a contradiction: $U \in M$ (because $U \in V$ and $V \in U$) but $U \notin \emptyset$), and $\emptyset \notin M$. $M \neq U$ because $\emptyset \notin M$. $M = V$ would be a contradiction: in that case M should be empty with the only possible exceptions being U and M itself, but if M contains U, it must do so because of some e distinct from both M and U such that $U \in e$ and $e \in U$, and then M would also contain this e. So we have $M \neq V$. But this also leads to a contradiction: $M \in W$ (from the assumption that $W = U$) but also $M \notin W$: because $\emptyset \notin M$ and $V \in M$ (this is because $M \neq V$ and $M \neq U$, and $U \in V$ and $V \in U$) so M does not satisfy the criterion in the definition of W.

So W is a new set (i.e. distinct from U, $\emptyset$, and V)! Now we can see that $\emptyset \in W$, $U \in W$, $W \in U$, $W \notin \emptyset$, and $V \notin W$ because $\emptyset \notin V$ and $U \in V$ so V does not satisfy the disjunctive criterion used in the definition of W. (We do not need to decide whether $W \in V$ for the rest of the proof, so this lemma is indifferent to it.) □

Lemma 4. *$L \neq \emptyset$ and $\emptyset \in L$*[5]*.*

Proof. Proving this lemma comes down to proving the inequality (then $\emptyset \in L$ follows from the definitions) and for this we use Lemma 3. If $L = \emptyset$, then $L \neq V$ and thus $V \notin L$. From the definition of L there must be some z distinct from L and V such that $z \in V$ and $V \in z$ and there is some y such that $y \neq z$, $y \neq V$, $y \neq L$, and $y \notin z$. But from the definition of V, every $z \in V$ is such that it contains everything but possibly z and V, so there can be no such y. This is a contradiction. □

[4] Note that in this case, similarly to many other cases, $U \in W$ is not yet established, only that this must be the case if W were identical to $\emptyset$. There is still the possibility that $U = W$ and $U \notin W$. One needs to be careful at every step!

[5] Indeed also $L \notin \emptyset$. I shall no longer emphasize that everything distinct from U is in U and nothing distinct from $\emptyset$ is in $\emptyset$.

To prove $L \neq U$ I will use two sets D, E, both distinct from U, that do not satisfy the criterion of L – they are loopy via each other and none of them is such that it contains all sets except possibly L and itself. We already have the definition of E whereas D_1 is just a candidate for D. I shall prove that if D_1 does not satisfy the requirements for D, there is some D_2 that does. Note that we will later need $E \neq L$ which is why there is the disjunct "$\forall y y \in x$" in the definition of E – to make sure that it contains U while (as will be shown later) L does not.

Lemma 5. *$E \neq U$, $U \in E$, $E \neq \emptyset$, $\emptyset \in E$, $V \neq E$ and $V \in E$, and $W \neq E$ and $W \notin E$, and there is some set D such that $W \neq D$, $D \neq U$, $D \neq E$, $D \in E$, $E \in D$, $D \neq \emptyset$, and $\emptyset \notin D$.*

Proof. Consider the sets E and D_1 defined above. $D_1 \neq U$ (otherwise there would be a contradiction: $\emptyset \in D_1$ (because $D_1 = U$) and $\emptyset \notin D_1$ (from the definition of D_1)). Then clearly $U \in D_1$. Also $E \neq U$ for the following reason. Suppose $E = U$. Then $W \in E$. But we know (from Lemma 3) that $\emptyset \in W$ and $V \notin W$, so W does not satisfy neither of the disjuncts in the criterion of E. Because $E \neq U$, clearly $U \in E$. Thus also $\emptyset \neq E$ and $\emptyset \in E$.

Also $D_1 \neq \emptyset$ (otherwise a contradiction: $U \in D_1$ but $U \notin \emptyset$) and $\emptyset \notin D_1$ from their definitions.

It must be that $E \neq D_1$ because $\emptyset \in E$ and $\emptyset \notin D_1$. Further we have $E \in D_1$ because $\emptyset \in E$. Now, if $D_1 \in E$, take this D_1 for D and the lemma will hold.

Suppose, on the other hand, that $D_1 \notin E$. From this and from the definition of E (note that the first disjunct in the definition of E is not true here because $\emptyset \notin D_1$) there must be some D_2 such that $D_2 \neq E$, $D_2 \neq D_1$, $D_2 \in D_1$, and D_2 does not contain anything else than possibly D_2, D_1, and E. Because $D_2 \in D_1$, then from the definition of D_1 it must be that $E \in D_2$. Also, $D_2 \in E$ because if D_2 has any member other than E and D_2, it is D_1, and the criterion of E is satisfied when D_2 substituted for x and D_1 for y – e.g. because $U \in D_1$. $D_2 \neq \emptyset$ (because $E \in D_2$) and $\emptyset \notin D_2$. Then also $D_2 \neq U$. So in case $D_1 \notin E$, take D_2 for D and the lemma will hold.

The last fact about D in this lemma to be proved is that $W \neq D$ which follows from $\emptyset \notin D$ and is thus true in both cases.

$V \neq E$ because $\emptyset \in E$ but $\emptyset \notin V$. $V \in E$ because every member of V (distinct from V itself) contains e.g. $\emptyset$.

Finally, $W \neq E$ because $V \notin W$ (from Lemma 3) while $V \in E$. $W \notin E$ because $\emptyset \in W$ and $V \notin W$ so W satisfies neither of the two disjuncts in the definition of E. □

Now that we have D and E we can prove $L \neq U$.

Lemma 6. *$L \neq U$, $U \notin L$, $L \neq W$, $L \neq E$ and $E \notin L$, $L \neq D$, and $D \notin L$.*

Proof. Suppose $L = U$. This together with Lemma 5 gives us a contradiction: $E \in L$ (because $L = U$) but also $E \notin L$ from the definition of L because E is loopy via D, both E and D are distinct from U and thus from L, they are also distinct from each other, and $\emptyset \notin D$, so E does not satisfy the criterion in the definition of L.

$V \neq L$ because $\emptyset \in L$ and $\emptyset \notin V$. Therefore $U \notin L$ because U is loopy via V and e.g. $\emptyset \notin V$ so U does not satisfy the criterion of L.

$L \neq W$ because $U \in W$ but $U \notin L$.

$L \neq E$ because $U \in E$ but $U \notin L$. $L \neq D$ because $\emptyset \in L$ and $\emptyset \notin D$.

$E \notin L$ because L, E, D are pairwise distinct, E is loopy via D, and D satisfies the last conjunct in the definition of L because $\emptyset \notin D$ (note that for this we also need $\emptyset \neq L$ from Lemma 4, and $\emptyset \neq D$ and $\emptyset \neq E$ from Lemma 5). Similarly, $D \notin L$ because D is loopy via E and E satisfies the last conjunct in the definition of L because $W \notin E$ (note that here we need $W \neq L$, and $W \neq E$ and $W \neq D$ from Lemma 5). □

Now it is finally time to turn to A and B, show that they are distinct from L, distinct from each other, and that a contradiction follows.

Lemma 7. *$A \neq U$, $U \in A$, $A \neq \emptyset$, $\emptyset \in A$, $A \neq L$, $D \neq A$, $D \notin A$.*

Proof. Suppose $A = U$. Then there is a contradiction: $D \in U$ but $D \notin A$ because $D \notin L$ and $\emptyset \notin D$ so D does not satisfy neither of the two disjuncts in the definition of A. Thus $A \neq U$. Then clearly $U \in A$. Also $A \neq \emptyset$ because $U \in A$. Thus $\emptyset \in A$ because $\emptyset \in L$. $A \neq L$ because $U \in A$ but $U \notin L$. $D \neq A$ because $\emptyset \in A$ but $\emptyset \notin D$. Then $D \notin A$ because $D \neq L$ and $D \notin L$ (from Lemma 6) and $\emptyset \notin D$ so D satisfies neither of the two disjuncts in the definition of A. □

Lemma 8. *$B \neq U$, $U \notin B$, $B \neq L$, and $B \neq A$.*

Proof. $B \neq U$ because there would be a contradiction: $\emptyset \in U$ but $\emptyset \notin B$ (from the definition of B). $U \notin B$ because $U \notin L$. $B \neq L$ because there would be a contradiction: $\emptyset \notin B$ but $\emptyset \in L$. $B \neq A$ because $U \in A$ but $U \notin B$. □

Lemma 9. *$B \in L$, $A \in L$.*

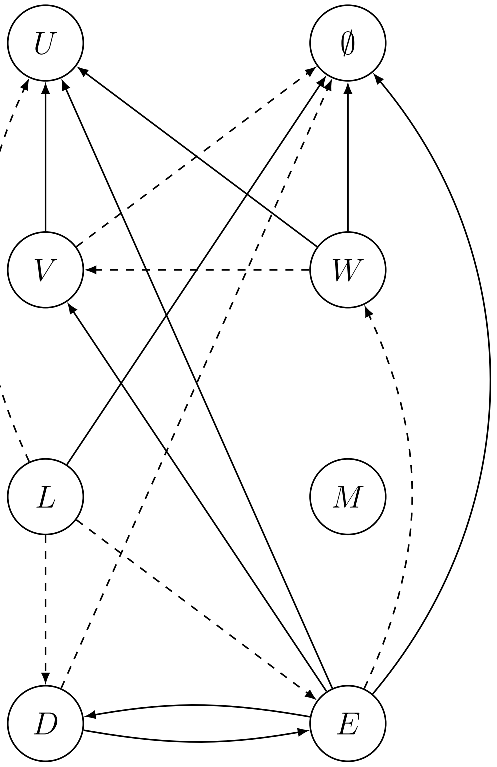
Proof. Suppose $B \notin L$. That means from the definition of L that there is some z such that $z \neq B$, $z \neq L$, $z \in B$, $B \in z$. Then we have $z \in L$ from the definition of B. But that is a contradiction with the definition of L because z is loopy via B and B satisfies the last conjunct in the definition of L (e.g. because $U \notin B$). Thus $B \in L$.

Suppose $A \notin L$. That means from the definition of L that there is some z such that $z \neq A$, $z \neq L$, $z \in A$, $A \in z$, and z does not contain some y distinct from A, L, and z. Then from the definition of A we have (i) $z \in L$ or (ii) z contains everything possibly except A, z, and L. (ii) is not possible because of the y. So it must be that $z \in L$. But z is loopy via A so $z \in L$ implies that A contains everything except possibly L, A and z, which is a contradiction with $D \notin A$ from Lemma 7 and the fact that D is distinct from L (Lemma 6), distinct from A (Lemma 7), and also distinct from z because $z \in A$ but $D \notin A$. □

Theorem 10. *The theory T_0 is inconsistent.*

Proof. From the previous lemmas we know that A, B, and L are pairwise distinct and we also have $A \in L$ and $B \in L$. Then we have (from the definition of B and the fact that e.g. $U \in A$) $A \in B$. We also have $B \in A$ because $B \in L$. But this is a contradiction: $A \notin L$ because A is loopy via B and e.g. $\emptyset \notin B$, so A does not meet the criterion in the definition of L. □

Figure 3.2: *Results of the lemmas 2–6 are depicted. Full arrows depict the relation 'contains' (inverse of $\in$), dashed arrows its complement. Only the results of the lemmas, not all the relations are depicted. In particular the full arrows from U to all other nodes and dashed arrows from $\emptyset$ are omitted.*



Note that the sets A and B are missing in Figure 3.2. All the depicted sets are defined without parameters. Therefore, if T_{WP} is consistent, the picture depicts the relations among these sets in every model. Adding parameters to comprehension adds the possibility of defining A and B and leads to a contradiction and it is therefore meaningless to talk about membership relations in this inconsistent theory (i.e. without any model).

Regarding the proof, note that the use of parameters is crucial. For example, we might define without parameters L as a set of non-loopy sets, A as a set of non-loopy sets and sets containing all the sets, and B as a set of non-loopy but also non-empty sets. Such a situation is crucially different from the situation where A and B are defined by L as a parameter. In the former, the property of being loopy used in L is ignoring loops via L, but the property of being loopy used in A is ignoring loops via A, and analogously for B. Therefore, as long as there is a set loopy only via A, like $\{A\}$, A will not be a member of L. And there is no paradox in that $A \in \{A\}$ and $\{A\} \in A$ because A is defined in such a way that loops via A are ignored. In contrast, if A is defined as in the proof by the parameter L, A cannot contain $\{A\}$ because L cannot not contain it.

3.2 Consequences of the inconsistency

3.2.1 Inconsistency of other exclusive set theories

In this section I show how the inconsistency of T_0 implies inconsistency of other exclusive set theories.

To analyse the alternatives, it is useful to look at the general form of comprehension:

$$\forall\overline{p}\exists S\forall x(x \in S \leftrightarrow \varphi(x,\overline{p}))$$

and divide the quantifiers in it into four groups:

1. parameters $\overline{p}$
2. $\exists S$
3. $\forall x$
4. quantifiers in φ

Note, however, that having the three kinds of quantifier – strongly exclusive denoted by the superscript 's', weakly exclusive by 'w', and inclusive by 'i' – is not sufficient to express all the options. Therefore, I will instead say e.g. "exclusivity 1-3", meaning that the bound variable in group 3 (i.e. $\forall x$) cannot have the value that is already taken by some variable quantified in group 1 (i.e. the value of a parameter). Note that this exclusivity 1-3, among others, cannot be signified by a formula using the three kinds of quantifier. For instance, if one writes:

$$\forall^i\overline{p}\exists^i S\forall^s x(x \in S \leftrightarrow \varphi^i(x,\overline{p})),$$

one signifies that the only exclusive quantifier is $\forall x$ (which is correct) but it is exclusive both with respect to the parameters and with respect to $\exists S$. So this would be exclusivity 1-3 together with 2-3[6].

While some combinations of choices of exclusivity would seem arbitrary, this section aims at a general result. Exclusivities which cannot be siginified by quantifiers are still meaningful and recall that the standard logic with only inclusive quantifiers is sufficient for formalising these exclusivities by simply putting inequalities to the appropriate places in the comprehension schema.

Recall that the inconsistent theory T_0 has an additional axiom saying that there are at least three sets in the universe. It is tacitly assumed below that the considered alternatives to T_0 entail this axiom. Clearly, set theories with only one or two sets are not of much interest.

Note that Frege's attempted way out of the paradoxes corresponds to having only exclusivity 2-3 (cf. [30]). This avoids Russell's paradox but almost nothing

[6] A sidenote: a similar problem, due to the strict linear order of quantifiers in formulae, exists even in classical logic with all quantifiers being inclusive. There is e.g. no way to have a formula like $\forall x\exists y\forall a\exists bF(x,y,a,b)$ except that the choice of y is only dependent on x and the choice of b is only dependent on a (the choice of b in the formula just given is also dependent on x). This is why Hintikka co-originated Independence friendly logic, see e.g. [31].

else: e.g. the paradox of non-loopy sets stays untouched. Then, Hintikka's first theory T_{HF} corresponds to exclusivity 2-3 and 2-4. As mentioned, this has been proved inconsistent already by Hintikka. T_{HS}, the inconsistency of which had been left open but follows from the result above, corresponds to exclusivities 2-3, 2-4, and 1-3. (Note that the exclusivity 1-3 means e.g. that we do not have standard singletons and pairs which prevents the paradox of T_{HF}.) Recall also that both T_{HF} and T_{HS} allow the use of identity in their comprehension schemas as well as all other theories considered by Hintikka.

The inconsistency of the whole family of exclusive set theories considered by Hintikka

T_0 corresponds to exclusivity between everything but 1-2 and does not allow identity in φ. I shall now argue that the inconsistency of T_0 implies inconsistency of every other alternative comprehension schema where identity is allowed. Firstly, it we cannot have inclusivity 2-3 or 2-4 because S must be excluded from all the subsequent quantifiers, otherwise Russell's VCP is violated. Accordingly, this is the case in both Hintikka's theories T_{HF} and T_{HS}. This is a minimal amount of exclusivity which Hintikka considered. Indeed, exclusivity 2-3 is needed for avoiding Russell's paradox, and exclusivity 2-4 needs to be added to avoid paradoxes like the one of non-loopy sets.

Suppose now that we have any theory T like T_0 except with an alternative choice of exclusivity vs inclusivity of quantifiers in its comprehension schema, and suppose T has at least exclusivity 2-3 and 2-4 and that identity is allowed in its comprehension. Such a theory entails T_0 for the following reason.

Let φ be any formula used in a comprehension in T_0. If any of the choices 1-4, 3-4, or 4-4 in T are inclusive, we can translate φ into a new formula ψ that will be exclusive in 1-4, 3-4, or 4-4 by using identity as described earlier in the thesis. This way we get ψ which is equivalent to φ, or more precisely, φ with exclusivity determined by T_0 is equivalent to ψ with exclusivity determined by T[7].

As the groups 2 and 3 only contain one quantifier (and thus exclusivities 2-2 and 3-3 are meaningless), all the remaining cases of a possible difference between T and T_0 are inclusivity vs exclusivity in 1-1, 1-2, and 1-3. Recall that T_0 has exclusivity 1-1, inclusivity 1-2, and exclusivity 1-3. It is clear that other options for 1-1, 1-2, and 1-3 are stronger than that of T_0: if something is true for *all* parameters p_1, ..., p_n, it is also true for *all pairwise distinct* p_1, ..., p_n. If, for a given choice of parameters, there is S *distinct from all the parameters* such that ..., then there is *some* S such that... And if something is true for *all* x, it is also true for all x *distinct from the parameters and from* S.

In short, given any comprehension axiom of T_0 with φ, one can translate this φ into ψ by adding any 1-4, 3-4, or 4-4 exclusivity which is not present in T "manually" using identity, and the comprehension axiom of T given by this ψ entails the axiom of T_0. Because any such theory T entails T_0, a consequence of the inconsistency of T_0 is inconsistency of all such theories – i.e., all theories

[7] Both this equivalence and the claim that T entails T_0 can be understood either semantically or syntactically. Semantically, the equivalence means that they are true in the same structures of the set-theoretical language. In that case, the whole argument shows that if T' had a model, T_0 would also have a model. But T_0 does not have a model, thus T does not either (i.e., it is inconsistent).

like T_0 except with a possibly different choices of exclusivity vs inclusivity in its comprehension schema and with identity allowed in the comprehension schema.

Hintikka's open problem answered

Hintikka [2] does not seem to consider theories without identity in the comprehension. Thus when he says:

> It may be pointed out that our resources are by no means exhausted by [T_{HF}]. In fact, [T_{HF}] was the outcome of one particular reinterpretation of the variables of [the comprehension schema with all quantifiers inclusive]. Other interpretations will lead to other systems, some of which are still safer than [T_{HF}]. For instance, there is a system based upon [T_{HS}][...] [2, p. 242]

and leaves open the question of consistency of other systems, this question has just been answered: all of these are also inconsistent.

One might want to consider theories without identity in their comprehension schemas and different from T_0 in at least one of the choices among 1-4, 3-4, and 4-4, because inconsistency of these does not straightforwardly follow from the inconsistency of T_0. If a theory T does not allow identity and has, unlike T_0, inclusivity 1-4, 3-4, or 4-4, it does not entail T_0 for the reasons sketched above because the translated formulae ψ's use identity. Note also that identity of sets $a = b$ cannot be paraphrased without identity by imitating the axiom of extensionality[8] because S cannot be mentioned in φ, thus two distinct sets might pass such a test if they only differ in that one contains S and the other does not. It would be surprising to me if one of these theories were consistent and it seems more likely that also in these theories parameters could be used to derive contradictions, but I have no proof of that as I have not considered this question in any detail.

3.2.2 What options are left?

If parameters in comprehension lead to inconsistency, one might give them up altogether. In that case, T_{WP} is the most natural choice. If T_{WP} is consistent, it is in some ways richer than the standard theories (e.g. it has a set of all sets (possibly except itself, of course), it also has Frege's numbers and other big sets) and even in some ways richer than New Foundations (e.g. it has a set L of all non-loopy sets, although this non-loopiness is relative to the set L itself). It is, however, very non-classical in the sense that e.g. singletons and pairs do not necessarily exist (as explained earlier in the thesis, it does not follow from the axioms that, e.g., for any set x there is the set $\{x\}$ with just one member – x). And if one does not have pairs, one does not (presumably) have ordered pairs and functions either. Consequently, is is not clear how a mathematician could work with such a theory. Perhaps it might be worth considering further, but I have no solution to the problems.

[8] By a formula like: $\forall z(z \in a \leftrightarrow z \in b) \wedge (a \in a \leftrightarrow a \in b) \wedge (b \in a \leftrightarrow b \in b) \wedge (x \in a \leftrightarrow x \in b) \wedge ... \wedge (p_1 \in a \leftrightarrow p_1 \in b) \wedge ... \wedge (p_n \in a \leftrightarrow p_n \in b)$.

Alternatively, one might consider adding other axioms, but recall that adding the axiom of pairing leads to inconsistency.

One might also change the axiom of extensionality to something more akin to its alternatives used in non-well-founded set theories (see [29]) and hope that this would avoid inconsistency of some exclusive set theory with parameters. However, this seems unlikely to work. The whole idea of using exclusive interpretation was to avoid circularity, but it seems that the circularity remains and it will remain regardless of the version of axiom of extensionality. Circularity seems to arise from the unrestricted comprehension schema. This is partly addressed in the next chapter.

4. Philosophical interpretation

Russell's VCP is often considered as too strong because it bans many impredicative definitions that are commonly used in mathematics and seem unproblematic, like the definition of the least upper bound. However, it seems that it is also in some cases too weak: the exclusive theories considered in Chapter 2 and Chapter 3 are in accordance with the principle, yet they are inconsistent. This chapter discusses why and how Russell's VCP fails and how it could potentially be improved.

4.1 Russell's VCP misses the mark

Russell's VCP, as formulated in 1.1.3, seems to miss the mark. In this section I mention three reasons for this.

The first reason is that the inconsistent exclusive theories seem to be in accordance with the principle. As mentioned in 1.1.3, Hintikka argued against Russell's VCP based on the fact that one natural interpretation of this principle leads to an inconsistent set theory – T_{HF}. This argument is made even stronger by the results in Chapter 3. Now we can say that *several* natural interpretations of Russell's VCP lead to an inconsistent theory.

The second reason is similar to the first but is concerned with paradoxes outside of set theory. Attempts to solve non-set-theoretical paradoxes by exclusivity also seem to be in accordance with Russell's VCP but also fail. Recall from Section 1.1.3 that Poincaré suggested such a solution to Richard's paradox. It could be similarly suggested to solve Berry's paradox by defining a as "the smallest natural number that is not definable in fifty syllables" and with exclusive interpretation of such a definition – in such a case the class of natural numbers not definable in fifty syllables excludes a (regardless of whether a has this property) and also excludes all numbers such that their definitions in fifty or less syllables in some way reference a. Although such a "solution" avoids Berry's paradox of a being both definable and not definable in fifty syllables, it is not sufficient. Consider for example the definitions (again with exclusive interpretations of these definitions) of b (c) as the smallest *even* (*odd*) natural number greater than all natural numbers definable in fifty syllables. Such numbers b and c must necessarily exist because there are only finitely many numbers with a definition of fifty or less syllables, thus the classes of all even (odd) numbers that do not have such a definition are non-empty, thus they have smallest elements – b (c). Now, because of their different parity, b must be distinct from c. The paradox is in that b is supposed to be greater than c because c is defined in fifty syllables, is distinct from b and its definition in no way references b nor its definition, but by analogous reasoning also c is supposed to be greater than b.

While the first two reasons for invalidity of Russell's VCP show that it is not sufficient to guard us from paradoxes, the third reason shows that it is not necessary to subscribe to Russell's VCP either. There are definitions which violate Russell's VCP but seem unproblematic. An often cited example is that of "the tallest man in the room" which seems to be unproblematic even though it violates Russell's VCP. Ramsey says:

> [W]e may refer to a man as the tallest in a group, thus identifying him by means of a totality of which he is himself a member without there being any vicious circle. [32, p. 368]

The example of "the tallest man" is just one among many and many others are used in mathematics. An often quoted mathematical example is that of the least upper bound of a set. Every non-empty set of real numbers which has an upper bound in real numbers also has the least upper bound in real numbers. The definition of the least upper bound violates Russell's VCP because it invokes the class of all upper bounds to which the least upper bound itself belongs. Although the most radical predicativists would see this definition as illegitimate, they are in a clear minority.

Consequently, it has been argued that in cases in which the defined objects exist independently of us, Russell's VCP is not valid (Gödel [10], Ramsey [32], Chihara [8]). For example, Chihara writes:

> From the point of view of those who think that there really are sets that exist independently of human thoughts and practices the vicious circle principle is false. [8, p. 42]

Recall also the quotations from Gödel in Section 1.1.3.

4.2 Russell's VCP is too strong

There seem to be two problems with Russell's VCP which are in turn addressed in this and the following section. One problem makes Russell's VCP too strong in some cases in that presumably unproblematic definitions violate it. The other makes Russell's VCP too weak in some cases in that it is insufficient to avoid circular paradoxes.

The reason why Russell's VCP is too strong in some cases can be explained on the following two examples. Consider the following two definitions of a real number which have already been mentioned. The first one is arguably legitimate while the other is illegitimate.

> Least upper bound: Given a set S of real numbers with at least one upper bound in real numbers, define a as the least upper bound of S.
>
> Richard's number: Given a countable ordered set S of definable real numbers, define b by diagonalisation as described in Section 1.1.3.

Both definitions violate Russell's VCP: the former defines a real number a by invoking the class of all upper bounds, to which a itself belongs; the latter purports to define a real number b by invoking the class of all definable real numbers, to which b itself should belong.

However, there is a crucial difference between these two definitions. In the latter case, and not in the former, also the definition itself is involved, not only the defined object.

To determine which real number is defined by the first definition, one should in principle go through all upper bounds of S, compare them and pick the least one. Or, presumably, one should go through all real numbers, determine which of them are upper bounds of S, and then determine which one is the least of these. In any case, one only goes through real numbers (i.e. objects) and not through definitions. Consequently, only the object in question is involved, not the definition in question. This is the case in so far as the definition in question is not somehow involved in "going through all real numbers", which it is not because real numbers (as well as upper bounds of S) can be defined independently of the definition in question.

However, consider now the latter definition – the definition of Richard's number. To determine which real number is defined by this definition, one should in principle go through all definable real numbers and use them in the construction. But going through all definable numbers means going through all the definitions of real numbers and looking at which objects are defined by them. Thus the definitions are involved, including the one in question – the definition of Richard's number.

Hopefully, comparing these two cases gives the reader some idea (albeit vague) what is crucial for the existence of vicious circles. It seems to be the involvement of something in the definition in question but this something is not the object which is being defined. What exactly is it, then?

It seems to me that a promising way to develop the idea more precisely would be to use Frege's distinction of sense and reference[1] [18]. Consider any case of a definition d which is supposed to define an object. For example, in the case of Richard's (Berry's) paradox we have a definition "The smallest real (natural) number such that..." which is supposed to define a real (natural) number. In the case of set theories, the definition is "the set of those sets which..." and it is supposed to define a set[2]. First of all, one can distinguish the syntactical and the semantical aspect of the definition – the former being the string of symbols used to represent the definition, the latter being the meaning. But a further question is: what is *the meaning* of a definition? Frege distinguishes between two "kinds of meaning" – *sense* and *reference*. The reference of a definition in question is the object defined by it, while the sense is described by Frege as follows:

> It is natural, now, to think of there being connected with a sign (name, combination of words, letter), besides that to which the sign refers, which may be called the referent of the sign, also what I would like to call the sense of the sign, wherein the mode of presentation is contained. [18, p. 210]

For example, consider the definition:

The third least prime number.

[1] The reader unfamiliar with this distinction is advised to familiarise themselves with it, although it might be possible to follow without doing so.

[2] One might perhaps use the term "singular term" instead of "definition". Clearly, I am concerned not with all kinds of definition in general but with the narrow case when the definition defined an object – be it a man, number, or set.

The syntax of this definition is the string "The third least prime number." Regarding its meaning, the reference of this definition is the natural number 5, and the sense of this definition could be approximated as *the mode of presentation* of the number 5 expressed by the definition.

The definition "the number which is the result of adding 2 to 3" would have the same reference as the definition above but a different sense.

Now it is possible to clarify what is meant by the claim that a definition d *may* involve the object which is defined by it, but it must not involve the definition d itself. Firstly, the syntax of the definition *may* be involved, consider for example the definition:

> This string of letters except with every 's' changed to 'z'.

This definition involves "its own syntax" but it seems unproblematic and does not give rise to a vicious circle. Its reference is clearly the string "Thiz ztring of letterz except with every 'z' changed to 'z'."

As argued above e.g. on an example of "the tallest man in the room", the reference may also be involved. Thus the following principle suggests itself:

> Sensical VCP (SVCP): A definition must not involve its own sense.

This is to be contrasted with:

> RVCP: The definition must not involve its own reference.

Return now to the the definition of Richard's number above. To determine which real numbers are definable, one must look at the definitions (including the definition of Richard's number) and determine which numbers are defined by them – i.e., what are their references.

Frege says the following about the connection of sense and reference in general:

> The regular connection between a sign, its sense, and its referent is of such a kind that to the sign there corresponds a definite sense and to that in turn a definite referent. [18, p. 211]

In the question of *which number is defined by a definition*, one starts with the sign of the definition, to this sign there corresponds a sense, and to this sense corresponds a reference – a definable number. In this way, senses of the definitions are involved, not only references.

Russell, in the context of VCP, does not seem to differentiate between a definition and the object defined by it. Without this distinction, SVCP and RVCP are indistinguishable. That Russell did not distinguish between the two can be seen e.g. in his formulation of VCP:

> "Whatever involves all of a collection must not be one of the collection." [14, p. 225]

What does this “Whatever” stand for, for the definition, or for the defined object? If for object, then it does not make sense because an object cannot be said to *involve* some object (as Gödel points out, recall the quotation in Section 1.1.3). If it stands for the definition, then this principle does not ban involving a collection including the denoted object, e.g. by means of quantifiers, but Russell *did* mean to ban involving the object. Thus it seems that Russell did not distinguish between a definition and the object it defines and “whatever” stands for both of them. Insisting on this distinction, the best approximation of this formulation of Russell seems to be:

> “Any definition that involves all of a collection must not be such that the object it defines is one of the collection.”

Recall also the formulation of VCP called “Russell’s VCP” in Section 1.1.3:

> “Everything that contains an apparent variable must be excluded from the possible values of this variable.”

Russell’s VCP seems to be a principle in the vein of RVCP. It says that the reference of the definition in question must be excluded from the possible values of the apparent variable – because objects (i.e. references of definitions) are values of variables, not their senses.

Consider again the definition of “The tallest man in the room.” This violates RVCP as the class of all men in the room is involved and the tallest man – the reference of this definition – is a member of this class. However, the sense of the definition is not involved in any way. The matter of *who the tallest man is* is settled completely by which men are in the room and how tall they are and the sense of the definition is in no way involved.

In summary, SVCP seems to improve RVCP in that it does not ban (in contrast to RVCP) the legitimate definitions like “the tallest man in the room” while it bans (just as RVCP does) those definitions that really do give rise to vicious circles like the ones in Berry’s and Richard’s paradox. Admittedly, the meaning of “involve” in SVCP is left quite vague. Whether a definition involves its own sense cannot be decided by a simple criterion.

4.3 Russell’s VCP is too weak

As argued in the previous section, RVCP is in general invalid because it blames the paradoxes on the involvement of the reference not of the sense of the definition. However, consider RVCP in the context of set theories discussed in this thesis. All these set theories (naive set theory, New Foundations, and the various exclusive set theories) have only the axiom of extensionality and the instances of the axiom schema of unrestricted comprehension as their axioms. Consequently, which sets exist is determined by the comprehension axioms. These comprehension axioms thus play a double role: on one hand they are used as definitions to single out a set from the totality of all sets; on the other hand they are used to determine which sets exist. Regarding the instances of the comprehension schema, every axiom saying that a set exists is at the same time a definition of this set, and vice versa.

Consider any definition like $S = \{x|\forall z(...)\}$. This definition, because of the quantifier, involves all sets z, including the set S itself, thus it violates RVCP. However, it also violates SVCP: what are *all sets*? That depends on which sets exist. And which sets exist? This depends on what sets are defined by comprehension axioms, i.e. by the definitions including this very definition of S. And as explained in the previous section in the example of Berry's paradox, the sense of a definition is involved in the quesiton of *what object is defined by this definition.* In this way, the sense of this definition of S is involved. For this reason, as long as the existence of sets is not determined independently of the definitions, SVCP is equivalent to RVCP. To separate them, one would need to describe what sets exist independently of the comprehension axioms. Then the *definition of the set* would depend on *all sets* which would depend on *which sets exist* which would depend on *whatever way we used to describe which sets exist* and not on *definitions of the sets.* This is perhaps the case of Zermelo-Fraenkel set theory.

Note that the paradoxes discussed in this thesis are often divided into two groups: *logical paradoxes* and *semantical paradoxes*. The former group includes set-theoretical paradoxes, the latter Berry's and Richard's. These two categories are thought to be different and deserving a different kind of a solution. On the other hand, Vicious circle principles (Russell's VCP, but also SVCP) can be seen as giving a solution to all the paradoxes, not recognising the categorisation. The categorisation, although being accepted by many, has its critics:

> Russell himself was unable to say what held the family of paradoxes together beyond some rather unsatisfactory remarks concerning vicious circles[...] It is therefore unsurprising that the modern view of the paradoxes is to the effect that there are two distinct families here, which arise from different sources, and which are to be treated quite differently. [...] [T]he founder of the orthodoxy was Ramsey (1925). [...] Russell was right and Ramsey was wrong. The paradoxes of self-reference do have a common underlying structure, which generates the contradiction involved[...] [33, p. 24]

Now, if SVCP is equivalent to RVCP in the context of set theories discussed in this thesis and if SVCP is valid, implementing RVCP in naive set theory should lead to a consistent theory. Exclusive set theories (for instance T_0) seem to implement RVCP (and thus SVCP) but are inconsistent. Why is this so?

This is because the inconsistent exclusive set theories only implement Russell's VCP in the narrow sense, where "involving" is narrowed to "involve by an apparent variable" in accordance with Russell's formulation mentioned in Section 1.1.3. However, "involving" must be understood in a broader sense.

We may consider several types of self-involvement, each one less direct than the previous one. Firstly, there is the most direct self-involvement by an immediate self-reference. Outside of set theory, this is the case of e.g. Liar's paradox ("This proposition is false."), where the proposition references itself[3]. In set theory this would be the case if S itself could be used in φ as a free variable. This is however forbidden in all the set theories including naive set theory, thus this

[3] And its sense is involved in determining the truth value, which is why SVCP is violated an there is a vicious circularity.

most direct involvement can be ignored in the context of set theories. A less direct involvement is via quantifiers – when a set S is defined by reference to *all* sets by a quantifier, one of the possible values of which it itself is. This involvement is recognised by Russell's VCP and is banned by requiring that the set being defined is excluded from the possible values of apparent variables. This involvement is also implemented by exclusive set theories. However, there are other, even less direct self-involvements.

In exclusive set theories, when a set S is defined using quantifiers, the possible values of these quantifiers exclude S itself. However, they include other sets and these other sets are defined by exclusive quantifiers excluding the values of themselves but not of S. Thus it so happens that S involves some set T (because T is a possible value of quantifiers used in the definition of S) and this T involves S (because S is a possible value of quantifiers used in the definition of T).

The most immediate self-involvement by using S itself in φ could be named 0-self-involvement, the one banned by Russell's VCP 1-self-involvement, and the one described in the paragraph above 2-self-involvement. This indeed generalises for even greater natural numbers than 2.

If one looks at the problem only in light of RVCP, it is impossible to see 2-self-involvement and higher self-involvements. This is because the reference is an object and an object cannot be said to involve something – only definitions or their senses do. But if one recognises both SVCP and RVCP and sees their equivalence, the matter becomes clear. In the case of 2-self-involvement, the definition of a set S involves some T. Because of the equivalence of SVCP and RVCP in set theories, S also involves the sense of the definition of T. The sense of T involves (by use of quantifiers) other sets, including S. Thus S eventually (via T) involves itself.

Note that this situation in set theories is analogous to the exclusive Berry's paradox described in the beginning of this chapter. There, too, the defined number b does not 0-involve nor 1-involve itself, but it 2-involves itself via c.

The last problem to be clarified is this. The analysis above would suggest that NF or T_{WP} also allow definitions which involve themselves, and thus these theories would also violate SVCP and RVCP. No use of parameters is needed for a definition of a set to 2-involve itself in the way described above, and the stratification requirement in NF does not prevent it either. So why are they not circular? The only possible answer is: they are circular, but maybe the circles are not vicious.

The self-involvement described above leads to circularity. A vicious circularity is different from circularity in that it leads to paradoxes.

The fact that circularity does not necessarily entail vicious circularity can perhaps be best show on a non-set-theoretical example. Consider Liar's paradox – the sentence "This proposition is false." This sentence 0-involves itself and is paradoxical. Suppose, however, that 0-self-involvement is banned and consider the following example.

Suppose Pinocchio utters the sentence s: "All sentences uttered by me during my whole life are false". Although s involves itself (its sense), it does not lead to a paradox if Pinocchio utters at least one true sentence during his lifetime – in that case, s is simply false. However, if he does not utter a true sentence, s is true if and only if it is false. Indeed, exclusivity does not help in this case either:

even when the sentences are interpreted exclusively, there is a circularity (in the form of 2-self-involvement) which can be vicious in some context. For example, suppose that Pinocchio only utters two sentences during his life: “All (*other*) sentences uttered by me during my life are false”, and “All (*other*) sentences uttered by me during my life are true”. This is a paradoxical situation.

The first example – with the sentence s – is analogous to a situation in New Foundations. Consider the Frege’s number two in NF: F_2 is the set of all sets with exactly two members.

$$\exists F_2 \forall x(x \in F_2 \leftrightarrow \exists a \exists b(a \neq b \wedge \forall c(c \in x \leftrightarrow (c = a \vee c = b)))).$$

F_2 involves itself but it clearly has more than two members. Suppose you should answer the question of whether $F_2 \in F_2$. You first determine e.g. that $\{\emptyset, \{\emptyset\}\}$, $\{\emptyset, \{\{\emptyset\}\}\}$, and $\{\{\emptyset\}, \{\{\emptyset\}\}\}$ are all members of F_2 and then answer the question: F_2 is not a member of F_2. This is analogous to the Pinocchio’s sentence s in the case he utters some true sentences during his life. Analogous to the case in which he does not utter a true sentence in his life would be a situation in which there were only two other sets than F_2 with two members. Then if $F_2 \notin F_2$, it has these two members, and thus it should be a member of itself. On the other hand, if $F_2 \in F_2$, it has three members and thus should not be a member of itself.

Indeed, there are infinitely many members of F_n for every positive natural number n in NF, so there is no such paradox in NF. But we can see that there is a circularity in NF which is not present in e.g. ZF.

In the consistent cases, it might be said that the definition of a set (or the statement by Pinocchio) *involves* itself yet it does not *depend on* itself.

The case with exclusive interpretation of Pinocchio’s sentences is analogous to the exclusive set theory T_{WP}. This theory is circular just like NF, naive set theory or other exclusive set theories (except there are no 1-self-involvements, only n-self-involvements for $n > 1$). Consequently, If T_{WP} is consistent, this is not because it is less circular than the inconsistent exclusive set theories like T_0. It is because this circularity is not vicious. The use of parameters in exclusive set theories does not add circularity but it turns circularity into vicious circularity. Because the set S being defined in T_{WP} is excluded from everything that can be said in the defining property, S arguably only *involves* itself but does not *depend on* itself – the property ignores it: if $\{\emptyset, S\}$ satisfies the property, then also $\{\emptyset\}$ does, etc. Parameters break this and turn circularity into vicious circularity (as mentioned below Theorem 10). The parameters change “involves” to “depends on”.

If NF or T_{WP} is consistent, it would seem more appropriate to rename “vicious circle principles” to “circularity principles”, at least as long as the existence of a vicious circle implies inconsistency. A valid circularity principle (SVCP is a candidate for such a principle) would guard us against circularity. Arguably, circularity is problematic and it is reasonable to avoid it regardless of whether it is vicious. However, in that case, if a theory violates the circularity principle, it is because it is circular and it might possibly be consistent.

There are indeed interesting question regarding circularity that have not been properly addressed, for example: *Is there a way to distinguish viciously circular theories from non-viciously circular?*, *Should circularity in itself be avoided or* *is it only problematic when it is vicious?*, and *Can the notion of "involving the sense" in SVCP be made more precise?*

Conclusion

In this thesis, I have followed up on Jaakko Hintikka's work on using exclusive interpretation of quantifiers to avoid paradoxes of naive set theory. Exclusive interpretation allows avoiding the well-know paradoxes of naive set theory while keeping its axiom of extensionality and axiom schema of unrestricted comprehension. Several such *exclusive set theories*, formalised in first-order logic, have been considered in this thesis. The primary criterion for success of an exclusive set theory is its consistency, the secondary criterion might be its usefulness. In this sense, the endeavour seems to be doomed to fail: allowing parameters in the comprehension schema leads to inconsistency while not allowing them seems to lead to an unworkable theory.

Hintikka left open the question of consistency of a family of exclusive set theories. The main contribution of this thesis is the proof that all these exclusive set theories are inconsistent. I have proved this by showing that a particular exclusive set theory which seems rather weak is inconsistent. From this, inconsistency of other exclusive set theories follows.

I have also discussed Russell's vicious circle principle, partly in light of inconsistency of exclusive set theories. I have argued that while Russell's vicious principle blocks a certain kind of circularity, it does not block other less direct kinds of circularity. This is why the exclusive set theories are inconsistent even though they do not violate the principle.

The problems with Russell's vicious circle principle and their connection to various paradoxes and to inconsistency of exclusive set theories offer directions for a further research. In particular, a way to improve Russell's principle is indicated in the last chapter. It seems worthwile to try and develop some ideas from the last chapter in more detail.

Bibliography

[1] Ludwig Wittgenstein. *Tractatus Logico-Philosophicus* (Pears and Mcguinness, Trans.; 2nd ed.). Routledge, 2001.

[2] Jaakko Hintikka. Identity, Variables, and Impredicative Definitions. *The Journal of Symbolic Logic*, 21(3):225–245, 1956.

[3] Kai F. Wehmeier. How to Live Without Identity–And Why. *Australasian Journal of Philosophy*, 90(4):761–777, 2012.

[4] Jaakko Hintikka. Vicious Circle Principle and the Paradoxes. *The Journal of Symbolic Logic*, 22(3):245–249, 1957.

[5] I. Grattan-Guinness. How Bertrand Russell discovered his paradox. *Historia mathematica*, 5(2):127–137, 1978.

[6] Peter Eldridge-Smith. In Search of Modal Hypodoxes Using Paradox Hypodox Duality. *Philosophia*, 50(5):2457–2476, 2022.

[7] Bertrand Russell. On Some Difficulties in the Theory of Transfinite Numbers and Order Types. *Proceedings of the London Mathematical Society*, 4(14):29–53, 1905.

[8] Charles S. Chihara. *Ontology and the Vicious Circle Principle.* Cornell University Press, 1973.

[9] W. V. Quine. New Foundations for Mathematical Logic. *The American Mathematical Monthly*, 44(2):70–80, 1937.

[10] Kurt Gödel. Russell's mathematical logic. In Paul Benacerraf and Hilary Putnam, editors, *Philosophy of Mathematics Selected Readings* (2nd ed.), pages 447–469. Cambridge University Press, 1983.

[11] Bertrand Russell. Some Explanations in Reply to Mr. Bradley. *Mind*, 19(75):373–378, 1910.

[12] Bertrand Russell. Les Paradoxes de la Logique. *Revue de Métaphysique et de Morale*, 14(5):627–650, 1906.

[13] Henri Poincaré. Les Mathématiques et la Logique. *Revue de Métaphysique et de Morale*, 14(3):294–317, 1906.

[14] Bertrand Russell. Mathematical Logic as Based on the Theory of Types. *American Journal of Mathematics*, 30(3):222–262, 1908.

[15] Henri Poincaré. La Logique de l'Infini. *Revue de Métaphysique et de Morale*, 17(4):461–482, 1909.

[16] Solomon Feferman. Systems of Predicative Analysis. *The Journal of Symbolic Logic*, 29(1):1–30, 1964.

[17] Solomon Feferman. Predicativity. In Stewart Shapiro, editor, *Oxford Handbook of Philosophy of Mathematics and Logic*, pages 590–624. Oxford University Press, 2005.

[18] Gottlob Frege. Sense and Reference. *The Philosophical Review*, 57(3):209–230, 1948.

[19] Bertrand Russell. *Principles of Mathematics.* Routledge, 1903.

[20] Marta Vlasáková. What is Identical? *Logica Universalis*, 15(2):153–170, 2021.

[21] Robert J. Fogelin. Wittgenstein on Identity. *Synthese*, 56(2):141–154, 1983.

[22] Robert Trueman. Eliminating Identity: A Reply to Wehmeier. *Australasian Journal of Philosophy*, 92(1):1–8, 2014.

[23] Kai F. Wehmeier. Still Living Without Identity: Reply to Trueman. *Australasian Journal of Philosophy*, 92(1):173–175, 2014.

[24] Thomas Forster. Quine's New Foundations. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy.* Metaphysics Research Lab, Stanford University, Summer 2019 edition, 2019.

[25] T. E. Forster. *Set Theory with a Universal Set: Exploring an Untyped Universe.* Clarendon Press, 1992.

[26] R. B. Jensen. On the Consistency of a Slight (?) Modification of Quine's "New Foundations". *Synthese*, 19(1/2):250–264, 1968.

[27] M. Randall Holmes. NF is Consistent, arXiv, 2022.

[28] W. V. Quine. On Cantor's Theorem. *The Journal of Symbolic Logic*, 2(3):120–124, 1937.

[29] Peter Aczel. *Non-Well-Founded Sets* (CSLI Lecture Notes: No. 14). CSLI Lecture Notes, 1988.

[30] W. V. Quine. On Frege's Way Out. *Mind*, 64(254):145–159, 1955.

[31] Jaakko Hintikka and Gabriel Sandu. Informational Independence as a Semantical Phenomenon. In Jens Erik Fenstad, Ivan T. Frolov, and Risto Hilpinen, editors, *Logic, Methodology and Philosophy of Science VIII*, volume 126 of *Studies in Logic and the Foundations of Mathematics*, pages 571–589. Elsevier, 1989.

[32] F. P. Ramsey. The Foundations of Mathematics. *Proceedings of the London Mathematical Society*, s2-25(1):338–384, 1926.

[33] Graham Priest. The Structure of the Paradoxes of Self-Reference. *Mind*, 103(409):25–34, 1994.

List of Figures

Appendix

Overview of various set theories mentioned in the thesis

All the theories mentioned in this thesis can be formalised in the standard first order logic with standard inclusive quantifiers. This is the approach taken here to make the theories as clearly comprehensible as possible for the reader.

Common features

All the theories have the following three feature in common:

> STANDARD LOGIC: The theory is formalised in the standard first-order logic *with identity*. This includes quantifiers: there are only standard inclusive quantifiers.
>
> SET-THEORETICAL LANGUAGE: The only non-logical symbol in the language of the theory is the binary predicate symbol '∈'.
>
> EXTENSIONALITY: The following is an axiom of the theory:
>
> $$\forall x \forall y (x = y \leftrightarrow \forall z (z \in x \leftrightarrow z \in y))$$
>
> .

Because these three features are common to all the considered theories, I put them together as a single composite feature:

> STANDARD SET-THEORETICAL LANGUAGE WITH EXTENSIONALITY: The theory has the features: STANDARD LOGIC, SET-THEORETICAL LANGUAGE, and EXTENSIONALITY.

Various comprehensions

All the theories have some kind of unrestricted comprehension schema.

Naive comprehension

> NAIVE COMPREHENSION: $\forall p_1 ... \forall p_n \exists s \forall x (x \in s \leftrightarrow \varphi(x, p_1, ..., p_n))$ is an axiom of the theory for every $n \in \mathbb{N}$ and for every formula φ with free variables x, p_1, ..., p_n.

Note that x, p_1, ..., p_n are the only variables occuring freely in φ. Thus, importantly, s does not appear as a free variable in φ.

NF comprehension

> NF COMPREHENSION: $\forall p_1 ... \forall p_n \exists s \forall x(x \in s \leftrightarrow \varphi(x, p_1, ..., p_n))$ is an axiom of the theory for every $n \in \mathbb{N}$ and for every *stratified* formula φ with free variables x, p_1, ..., p_n.

Where a formula φ is *stratified* iff there is an initial segment $S = \{0, 1, ..., k\}$ of natural numbers and a function σ from the set of all variables in φ to S such that:

(i) For every atomic formula $x = y$, we have $\sigma(x) = \sigma(y)$.

(ii) For every atomic formula $x \in y$, we have $\sigma(y) = \sigma(x) + 1$.

Note that there *is* the set $\{a, \{a\}\}$ for any set a because a can be the value of two distinct parameters in the comprehension schema. However, it is not possible to say that $x = \{y, \{y\}\}$ for some y by a stratified formula $\varphi(x)$.

T_{HF} comprehension

> T_{HF} COMPREHENSION:
>
> $$\forall p_1 ... \forall p_n \exists s \forall x(x \neq s \rightarrow (x \in s \leftrightarrow \varphi^{-s}(x, p_1, ..., p_n)))$$
>
> is an axiom of the theory for every $n \in \mathbb{N}$ and for every formula φ with free variables x, p_1, ..., p_n.

Where φ^{-s} is obtained from φ by replacing every quantifier $\exists y(...)$ by $\exists y(y \neq s \wedge ...)$ and every quantifier $\forall y(...)$ by $\forall y(y \neq s \rightarrow ...)$.

T_{HS} comprehension

> T_{HS} COMPREHENSION: $\forall p_1 ... \forall p_n \exists s \forall x((x \neq s \wedge x \neq p_1 \wedge ... \wedge x \neq p_n) \rightarrow (x \in s \leftrightarrow \varphi^{-s}(x, p_1, ..., p_n)))$ is an axiom of the theory for every $n \in \mathbb{N}$ and for every formula φ with free variables x, p_1, ..., p_n.

Where φ^{-s} is obtained from φ by replacing every quantifier $\exists y(...)$ by $\exists y(y \neq s \wedge ...)$ and every quantifier $\forall y(...)$ by $\forall y(y \neq s \rightarrow ...)$.

T_0 comprehension

> T_0 COMPREHENSION: Comprehension axioms of the theory are given by transforming every instance of the axiom schema *Naive comprehension without identity in* φ in the following way. Given an instance of Naive comprehension, rewrite sequentially (from left to write) every quantifier *except* $\exists s$ of the form $...\exists z(...)$ to $...\exists z(z \neq y_1 \wedge z \neq y_2 \wedge ...z \neq y_k \wedge ...)$ and every quantifier $...\forall z(...)$ to $...\forall z((z \neq y_1 \wedge z \neq y_2 \wedge ...z \neq y_k) \rightarrow ...)$ where y_1, ..., y_k are all the quantifiers in whose scope $\forall z$ appears.

For instance, consider the following instance of NAIVE COMPREHENSION:

$\forall p \exists s \forall x(x \in s \leftrightarrow (x \in p \vee \neg \exists z z \in x))$

Going from left to right, one first comes across $\forall p$. Because this is the first quantifier, it is not in the scope of any quantifier, thus it stays the same. Then one comes across $\exists s$ which is the only exception to the rewriting rule and is thus left the same. Then one comes across $\forall x$ and rewrites it by the rule given above to get the intermediate formula $\psi_1 \equiv \forall p \exists s \forall x((x \neq p \wedge x \neq s) \rightarrow (x \in s \leftrightarrow (x \in p \vee \neg \exists z z \in x)))$. The next and the last quantifier is $\exists z$ and rewriting it leads to the final formula $\psi_2 \equiv \forall p \exists s \forall x((x \neq p \wedge x \neq s) \rightarrow (x \in s \leftrightarrow (x \in p \vee \neg \exists z(z \neq p \wedge z \neq s \wedge z \neq x \wedge z \in x))))$. So this psi_2 is an axiom of T_0.

T_{WP} comprehension

> T_{WP} COMPREHENSION: Comprehension axioms of this theory are given in the same way as in the case of T_0 COMPREHENSION, except that only formulae $\varphi(x)$ *without parameters* are allowed.

Various set theories

Naive theory is given by the following features:

- STANDARD SET-THEORETICAL LANGUAGE WITH EXTENSIONALITY
- NAIVE COMPREHENSION

New Foundations is given by the following features:

- STANDARD SET-THEORETICAL LANGUAGE WITH EXTENSIONALITY
- NF COMPREHENSION

Hintikka's first theory T_{HF} is given by the following features:

- STANDARD SET-THEORETICAL LANGUAGE WITH EXTENSIONALITY
- T_{HF} COMPREHENSION

Hintikka's second theory T_{HS} is given by the following features:

- STANDARD SET-THEORETICAL LANGUAGE WITH EXTENSIONALITY
- T_{HS} COMPREHENSION

T_0 is the theory given by the following features:

- STANDARD SET-THEORETICAL LANGUAGE WITH EXTENSIONALITY
- T_0 COMPREHENSION
- ADDITIONAL AXIOM: $\exists x \exists y \exists z(x \neq y \wedge x \neq z \wedge y \neq z)$

T_{WP} is the theory given by the following features:

- STANDARD SET-THEORETICAL LANGUAGE WITH EXTENSIONALITY
- T_{WP} COMPREHENSION
- ADDITIONAL AXIOM: $\exists x \exists y \exists z(x \neq y \wedge x \neq z \wedge y \neq z)$