FORMAL FRAMEWORKS FOR CIRCULAR PHENOMENA

Possibilities of Modeling Pathological Expressions in Formal and Natural Languages

Kai-Uwe Kühnberger

Philosophische Dissertation
angenommen von der Neuphilologischen Fakultät
der Universität Tübingen
am 20.07.2001

Osnabrück
2002

Gedruckt mit Genehmigung der Neuphilologischen Fakultät der Universität Tübingen

Hauptberichterstatter:	Prof. Dr. Uwe Mönnich
Zweiter Berichterstatter:	PD Dr. Friedrich Hamm
Dritter Berichterstatter:	Prof. Dr. Schroeder-Heister
Vierter Berichterstatter:	Prof. Dr. Dr. Walter Hoering
Dekan:	Prof. Dr. Tilman Berger

Preface

Many times I thought that there will never be an end of this work. Finally, I can present this text. The seemingly infinitely many reasons for a delay of the final version of this dissertation can be categorized in a variety of aspects. I will not try to enumerate all these aspects, but I can say that there were not only technical problems that caused a further re-examination of several parts. Moreover there were private reasons, sometimes the lack of motivation, and last but not least changes in my environment.

In every preface of a doctoral thesis, people are mentioned who supported the dissertation and who helped fixing occurring problems. In this respect, the present dissertation is not an exception. These people do not only play an important role because of their scientific advice, but also because of their mental support. Sometimes this seems even more important than a sketch of the particular proof of a theorem. People who supported this work can be divided into two major groups: one group is physically (mainly) located in Tübingen, whereas the other one is physically (mainly) located in Bloomington, Indiana. I begin with the people from Tübingen. Many thanks to Prof. U. Mönnich for his help after my return to Germany in the summer of 1999, for offering me a job at the SFB 340 to finish this work, and for teaching me further aspects of coalgebras that are not related to circular phenomena. I am indebted to PD Fritz Hamm who taught me formal semantics and several applications of logical methods in linguistics. I would like to thank the present and former members of the Theoretical Computational Linguistics group in Tübingen, especially Dr. F. Morawietz, Dr. S. Kepser, Dr. J. Michaelis, and Dr. T. Cornell for the friendly and stimulating environment. Thanks to Prof. P. Schroeder-Heister and Prof. W. Hoering for the initial help to start the project at all.

Concerning the second group of people who supported this dissertation (namely the people in Bloomington, Indiana) I want to express my deep gratitude to Prof. L. Moss who was the father of many ideas and new directions in this dissertation. His research in various fields can be seen as the most important one for this work. Thanks to Prof. A. Gupta for teaching me revision theories and to Prof. D. McCarty and Prof. N. Cocchiarella for their support in Bloomington. As a matter of fact Prof. J. Barwise died in March 2000. His workshop at Indiana University, Bloomington in Spring 1998 was the initial motivation to learn channel theory. My deep respect to one of the most important and influential logician in the last decades. Finally I want to thank the Logic Group in Bloomington and numerous students for helpful discussions. Last but not least, I want to thank T. Crane for helping me to improve the English

of this dissertation. Nevertheless, I am responsible for my own concerning all mistakes and stylistic desiderata of this work.

The 'Graduiertenförderung' of Baden-Württemberg supported this dissertation financially while the author was working in Tübingen during the academic year 1996/1997. For the financial support during a two-years stay at Indiana University, Bloomington I would like to thank the DAAD (Deutscher Akademischer Austauschdienst).

Osnabrück, February 2002
K.-U. K.

Contents

Part I

Introduction

Chapter 1

Use of Symbols

The usage of mathematical notions and concepts in this work is in general similar to the usage that is common in mathematical discourses. Notions that are specific for a certain field and not generally known will be introduced and explained in the text. Some additional remarks concerning basic mathematical concepts are nevertheless necessary here, in order to guarantee an easy understanding of the different topics.

The logical notions are used as usual. The symbols $\forall, \exists, \neg, \wedge, \vee, \rightarrow$ denote universal quantification, existential quantification, negation, conjunction, disjunction, and implication, respectively. Propositional variables (atomic propositions) of the propositional calculus are denoted by small Latin letters like p, q, r etc. Variables of (first-order) predicate calculus are denoted by x, y, z etc. Truth values are denoted - as long as not notified otherwise - by T (truth) and F (false). In the case we shall work in non-classical logical systems that include more than two truth values (compare especially Part II of this work), we will specify the denotations of non-classical truth values (for example, *neither true nor false* or *both true and false*) in the text.

Set theoretic notions are the ones that are commonly used in set theory. The expression $x \in X$ means that x is an element of a set X. The symbol $\in$ is considered as an elementary notion of set theory. The standard axiomatization of set theory uses only $\in$ and logical symbols. The expression $X \subseteq Y$ denotes the subset relation in set theory and is defined as follows:

$$X \subseteq Y \Leftrightarrow \forall x : (x \in X \rightarrow x \in Y)$$

Similarly, $X \subset Y$ means that X is a proper subset of Y (i.e. $X \subseteq Y$ but $X \neq Y$). The empty set is denoted by $\emptyset$, and the notion of set theoretical complement (relative to a given set or class) is defined as usual:

$$X^C = \{x \mid x \notin X\}$$

The Cartesian product of two sets X and Y is denoted by $X \times Y$. We code this product of two given sets X and Y set theoretically as usual:[1] Formally,

[1] Notice that the pair $\langle x, y \rangle$ is standardly set theoretically coded by the expression

we represent this as follows:

$$X \times Y = \{\langle x, y\rangle : x \in X \wedge y \in Y\}$$

The power set operation $\wp$ applied to a given set X denotes the following collection:

$$\wp(X) = \{x \mid x \subseteq X\}$$

The set theoretic operations union, intersection, and difference of two sets are defined as usual:

$$X \cup Y = \{a \mid a \in X \vee a \in Y\}$$

$$\bigcup X = \{a \mid \exists y \in X : a \in y\}$$

$$X \cap Y = \{a \mid a \in X \wedge a \in Y\}$$

$$\bigcap X = \{a \mid \forall y \in X : a \in y\}$$

$$A - B = \{x \mid x \in A \wedge x \notin B\}$$

We will introduce quite explicitly basic ideas of classical set theory (ZFC) in Part IV of this work, in order to ensure that a sufficient foundation is provided for readers that are not familiar with axiomatic set theory.

Functions are denoted according to the usual conventions: $f : X \longrightarrow Y : x \longmapsto f(x)$, denotes a function f that maps a set X into a set Y such that every $x \in X$ is associated with a unique value $f(x)$. We will not require every time that f is defined totally on X. Functions are set theoretically considered as a uniquely specified set of pairs. More precisely, a function $f : X \longrightarrow Y$ can be associated with the following set A:

$$A = \{\langle x, y\rangle : x \in X \wedge y \in Y \wedge (\langle x, y_1\rangle \wedge \langle x, y_2\rangle) \rightarrow y_1 = y_2\}$$

Quite similar considerations hold for relations. The standard representation in set theory is the same as in the case of functions, namely a set of pairs for which the relation applies. The only difference is that we do not require the uniqueness condition for pairs: the expression

$$(\langle x, y_1\rangle \wedge \langle x, y_2\rangle) \rightarrow y_1 = y_2$$

is in general not a valid restriction in the case of arbitrary relations. Other concepts of standard axiomatic set theory (if needed for further consideration) will be introduced in the text. We refer the reader to Chapter 10 for a more explicit introduction into the topic.

Properties of functions like injectivity, surjectivity, bijectivity are defined as usual. A function $f : X \longrightarrow Y$ is called injective, if it holds:

$\{\{x\}, \{x, y\}\}$. The expression $X \times Y$ denotes therefore the collection of all pairs of the form $\{\{x\}, \{x, y\}\}$.

$$\forall x \forall y : f(x) = f(y) \rightarrow x = y$$

A function $f : X \longrightarrow Y$ is called surjective, if it holds:

$$\forall y \in Y \exists x \in X : f(x) = y$$

Finally, a function $f : X \longrightarrow Y$ is called bijective if f is both injective and surjective. Other properties of functions, in particular properties of functions in the context of the theory of partial orders and category theory are: Given a partial order $\mathbf{D} = \langle D, \leq \rangle$, a function $f : X \longrightarrow Y$ is called monotone if it holds:

$$\forall x \forall y : x \leq y \rightarrow f(x) \leq f(y)$$

Furthermore: If two structures $\mathbf{S} = \langle S, \oplus_1, \oplus_2, \dots, \oplus_n \rangle$ and $\mathbf{S}' = \langle S', \otimes_1, \otimes_2, \dots, \otimes_n \rangle$ are given, then a function $f : S \longrightarrow S'$ is called a homomorphism if it holds:

$$f(x \oplus_i y) = f(x) \otimes_i f(y)$$

Homomorphic images will be important when we introduce basic category theoretic terms. Given structures $\mathbf{S}$ and $\mathbf{S}'$ as above, a function $f : S \longrightarrow S'$ is called isomorphic, if f is a homomorphism and additionally bijective. We shall define the concepts monomorphism, endomorphism, epimorphism, isomorphism and automorphism. Assume two structures $\mathbf{S}$ and $\mathbf{S}'$ are given as above. A monomorphism $f : S \longrightarrow S'$ is an injective homomorphism. An endomorphism $f : S \longrightarrow S'$ is a homomorphism such that it holds $S = S'$. An epimorphism $f : S \longrightarrow S'$ is a surjective homomorphism. An isomorphism is a bijective homomorphism. An automorphism is a bijective homomorphism such that it holds $S = S'$. Generalizations from properties of functions to properties of functionals and further to properties of functors will be introduced in the text as well.

In Part III we will need some facts from standard recursion theory.[2] Most definitions and properties will be provided in the text, but here are some remarks and facts concerning some general topics of recursion theory. The classes of functions in the Baire-space $\mathcal{N} = \{f \mid f : \mathbb{N} \longrightarrow \mathbb{N}\} = \mathbb{N}^{\mathbb{N}}$ can be categorized according to the complexity of the elements of this class. Most important in this respect are the arithmetical and the analytical hierarchy. The arithmetical hierarchy can be characterized as follows. The complexity classes of functions are defined by the complexity of the definitions necessary to define the class. Whereas recursive functions can be defined without quantifying over variables, dependent on the complexity of the prefixes, the complexity increases. If the defining formula is of the form

[2] Still the best accessible textbook for recursion theory is [Ro67]. A readable introduction into recursion theory can be found in [Ob93]. Other resources for recursion theory are [So87] and [Od89]. For the structure theory of the analytic hierarchy, good textbooks are [Mo80] and [Ke95].

$$\exists x_1 \forall x_2 \exists x_3 ... Q x_n : R(x_1, x_2, ..., x_n)$$

where R is quantifier free and $Q \in \{\forall, \exists\}$, function f is in the complexity class Σ_n^0. In the case that f is definable with a formula of the form

$$\forall x_1 \exists x_2 \forall x_3 ... Q x_n : R(x_1, x_2, ..., x_n)$$

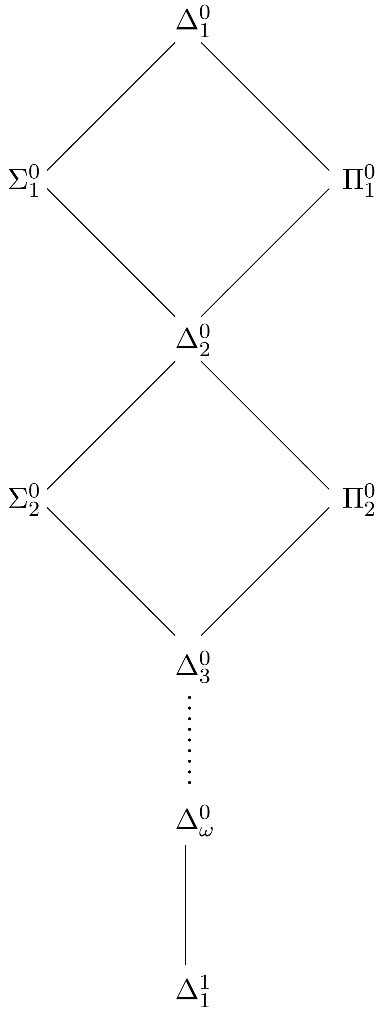
f is in complexity class Π_n^0. On the other hand, if f is definable via the formula

$$\forall x_1 \exists x_2 \forall x_3 ... Q x_n : R(x_1, x_2, ..., x_n)$$

and at the same time via the formula

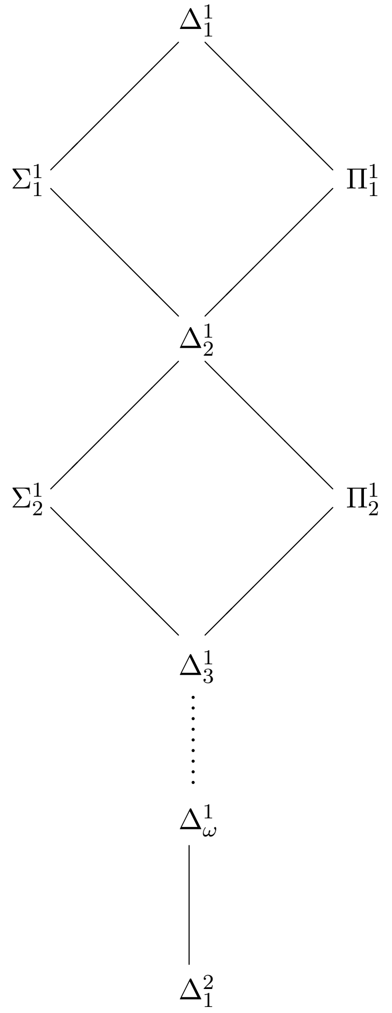
$$\exists x_1 \forall x_2 \exists x_3 ... Q x_n : R(x_1, x_2, ..., x_n)$$

then f is in the complexity class Δ_n^0 . We can represent this situation using a diagram (usually called the Kleene-hierarchy or the arithmetical hierarchy).



The complexity class Δ_1^0 corresponds to the recursive functions. From the perspective of descriptive set theory we can interpret Δ_1^0 sets as clopen (closed *and* open) sets.[3] The class Σ_1^0 corresponds to the class of recursively enumerable functions, whereas Π_1^0 corresponds to the semi-recursive functions. The arithmetical hierarchy represents the complexity classes of definable functions with a finite prefix of quantifiers ranging over variables.

In the arithmetical hierarchy, it is not possible to define functions via quantification over relations and functions themselves because quantification is restricted to individual variables. The definitional complexity of quantifications over functions and relations are ordered in the so-called analytical hierarchy (sometimes called projective hierarchy). The following diagram represents the analytical hierarchy.



In a certain sense, the idea behind the analytical hierarchy is the same

[3] We call a set X open, if it is closed under arbitrary unions and finite intersections. The collection of all closed sets can be identified with the class Σ_1^0. The complement of an open set is called closed. All closed sets can be identified with Π_1^0 sets.

as behind the arithmetical hierarchy. Whereas, in the arithmetical hierarchy quantification was restricted to the variables of the domain (usually numbers), in the more complex analytical hierarchy we can quantify over relations and functions as well. The Δ^1_1 functions are definable by both, a formula with one existential quantifier quantifying over relations and a formula with one universal quantifier quantifying over relations.[4]

There are several closure properties of the complexity classes that are important to mention (implicitly they are contained in the diagram above). For example, one can prove that the Π^1_n class (as well as the Σ^1_n class) really contains more sets than the Δ^1_n class. (The same is true for the analog complexity classes in the arithemtical hierarchy.) Δ^1_n sets are closed under negation, whereas this does not hold in general for Σ^1_n sets and Π^1_n sets. One can prove that it holds $\Sigma^1_n \cup \Pi^1_n \subseteq \Delta^1_{n+1}$, but it does not hold in general that $\Sigma^1_n \cup \Pi^1_n = \Delta^1_{n+1}$. The same holds in the arithmetical hierarchy: We have $\Sigma^0_n \cup \Pi^0_n \subseteq \Delta^0_{n+1}$ but in general it holds: $\Sigma^0_n \cup \Pi^0_n \neq \Delta^0_{n+1}$. It is important to notice that it does not hold in general: $\bigcup_{n>0} \Delta^0_n = \Delta^1_1$.

We are not trying to develop a complete picture of recursion theory here. In the case we need more information concerning recursion theoretic results and facts, we will provide them in the text.

This work will be partly based on algebraic representations of certain logics, semantical systems and non-well-founded set theory. Additionally, we will introduce basic ideas of the theory of partial orders, and of category theory. Concerning partial orders we use the following notions: $\langle X, \leq \rangle$ denotes a set X that is partially ordered by the relation $\leq$, i.e. $\leq$ is a relation that is reflexive, transitive and antisymmetric. The expression $\sup\{x, y\}$ denotes the supremum of the elements x and y where x and y are taken from the given set X. The supremum $\sup\{x, y\}$ of x and y is defined as the smallest upper bound of x and y. The notion $\sup(X)$ denotes the supremum of the set X. Sometimes, we will use an additional index: $\sup_D\{x, y\}$ denotes the supremum of x and y in the partially ordered set $\mathbf{D} = \langle D, \leq \rangle$. An alternative of $\sup\{x, y\}$ is to write $x \vee y$. We will use this shortcut quite often throughout the text.[5] If we denote the supremum of a set X we will often write $\bigvee(X)$. The dual of the operation supremum, namely the infimum of a given set X, is defined in a similar way. For example, $\inf\{x, y\}$ denotes the infimum of x and y, i.e. the largest lower bound of x and y. We will examine partially ordered sets in Part II, namely so-called bilattices. These structures are lattice-like structures, more precisely complete lattices ordered by using two order relations $\leq_1$ and $\leq_2$. Lattices are partially ordered sets $\mathbf{D} = \langle D, \leq \rangle$, such that for each two points x and y, $\sup\{x, y\}$ and $\inf\{x, y\}$ exist. A lattice $\mathbf{D} = \langle D, \leq \rangle$ is called complete if $\bigvee X$ and $\bigwedge X$ exists for every subset $X \subseteq D$. If $\mathbf{D}$ has a minimal element, this is in general denoted by $\perp$, and if $\mathbf{D}$ has a maximal element, this is denoted

[4] The hyperarithmetical sets (that are the Δ^1_1 sets) were relatively important for the development of recursion theory. For further information concerning recursion theoretic techniques the reader is referred to [Sa90, Od89], and [Ro67].

[5] A confusion with logical connectives is impossible when we use $\wedge$ instead the ordinary sup operation.

by $\top$. In special cases, we will use other symbols, but this will become clear in the context. Other concepts in the theory of partially ordered sets will be introduced in the text.

Concerning category theory, we will use a whole chapter (compare Chapter 12) in order to introduce the basic ideas, definitions, and results. The reader is referred to that chapter in order to get an overview of the concepts we use in this work.

A final remark concerning citations: In this work, reference to a particular Theorem, Lemma, Example etc. is denoted as usual: For example, Theorem 5.1.3 refers to the third numbered item stated in Chapter 5, Section 1.

Chapter 2

Phenomena

2.1 General Remarks

We begin our considerations with a discussion of phenomena that are altogether related to circularity in one or the other sense and are chosen from many different areas of scientific disciplines. There is a certain variety of circular phenomena and not all occurrences involving circularity result in a pathological behavior of the phenomena in question. First, we will examine circular sentences and discourses (that represent an appropriate context) in natural language that are at least in one respect reflexive. Essentially, we will deal in this part with the Liar paradox and related pathological sentences. Then, we will try to figure out which circular phenomena can arise in areas like knowledge representation and human reasoning in general. In this part, remarks about circular arguments are an additional aspect. In the third part of this chapter, we will examine circularity in other fields of academic disciplines like in linguistics, mathematics, and philosophy.[1]

When dealing with these topics, we are confronted with a principle question: Why is it important to know the behavior of pathological sentences at all? Why should we be able to analyze paradoxical expressions? More basically, we can ask the question, why is it necessary to model circular expressions in a given natural language with formal methods, if they are nevertheless somehow pathological? More drastically, one can ask: Why do we invest work and time for this problem at all? Some people argue that it is not necessary to analyze circular phenomena. Those people represent a totalitarian attitude towards scientific disciplines, claiming that someone knows which topics are worth considering.

There are practical reasons why we should try to examine models for circular phenomena. One reason is that only a formal model will enable us to program a machine that can deal and reason with reflexive phenomena. A second reason is that in order to understand language, we need a treatment of circularity because circularity is an aspect of natural language. Therefore, linguistic theories as well as the philosophy of language is incomplete until there is a developed theory of circularity. Furthermore, from a global perspective, the entire world provides us

[1] Compare [Ho79] for an exciting overview of various fields and disciplines in which circularity play an important role.

with a remarkable number of different kinds of circular phenomena. Applying logical methods is the attempt to find a better understanding and grasping of these phenomena by analyzing them in a formal framework. And in this sense, the examination of circularity is at the heart of scientific investigation.

In the following, we will give a variety of examples for circularity. Without claiming that these examples are in any sense complete, they will give an overview of the variety of the phenomenon circularity. And this variety can count as a certain motivation for our task. At least this can be taken as an argument for the necessity to consider this topic more closely. We will begin with some examples taken from natural language. First, we notice a fact. Natural language is strong enough to express sentences like (1):

(1) Everything the King says is true.

To try to give an English sentence that has the very same truth conditions as (1), but does not use the concept of truth fails practically. Even under the assumption that the following schema (the famous *convention T* introduced by Alfred Tarski)

(2) 'ϕ' is true iff ϕ (holds)

is an acceptable description of the properties of the truth predicate, we are faced with the practical problem to find a coextensional equivalent English sentence without using the truth concept. What happens if we try to eliminate the truth predicate in (1)? For a reliable representation of (1) without using the truth predicate we are forced to enumerate all sentences uttered by the King. Then, using convention T the conjunction of the list (or enumeration) should be (extensionally) equivalent to (1). It is hard to imagine that this is possible (at least from a practical point of view). Therefore, we can establish the conjecture that a language L which is equal to a language L^+ except for the lack of a truth concept is a sub-language of L^+.[2] In other words: a truth predicate in a language strengthens that language (at least in a practical respect).[3]

Another line of argumentation would be to notice that a precondition for the possibility of an artificially intelligent machine which is able to process natural language is that this machine can process sentences that include the truth predicate. This cannot be done in an appropriate way without a formal

[2] Someone could argue that it is possible to enumerate all expressions of the King (because the set of all utterances made by the King is finite), and therefore we have two equivalent languages, one with and one without a truth concept. Then, consider a sentence of the form: 'All theorems of Peano-Arithmetic are true'. Then, we are no longer able to argue against the claim that L is a sub-language of L^+ because the set of all theorems of Peano-Arithmetic is infinite and the introduction of infinite conjunctions is a proper extension of the original language L.

[3] We argue in this context from an intuitive point of view in order to give a flavor of the topic.

treatment or modeling of truth predicates in natural languages. In this respect, the practical applications make it necessary to develop a theory for the truth predicate.

The importance of truth predicates with respect to circularity can be seen by the fact that archetypical examples of circular sentences (of natural language) are usually constructed using the truth predicate. The results are usually pathological sentences. Because of this connection between a theory of truth and the strength of natural language to construct reflexive sentences, we reach a point where we can begin our considerations.

In the following section, I shall summarize different types of self-referential or circular expressions in English. I do not think that it is possible to give a complete list of such phenomena, so that the following is in a certain sense a fragment of different occurrences of the problem. Nevertheless, it is possible to give a large overview of the problems occurring in natural language because of circularity.

2.2 Truth and Paradox

The different forms of paradoxical sentences shall be analyzed in a pre-theoretical manner, with an approximation of the understanding of such sentences within naive reasoning. In most cases, intuitions, whether a particular sentence (in general a given 'statement' or 'utterance') is true or false, depend on ordinary (intuitive) reasoning. Clearly, this reasoning is in most cases close to (or equal to) classical two-valued logic. For example, we think that a sentence like

(3) John is playing the piano and Claire is buttering a toast.

is true if and only if both conjuncts are true, i.e. 'John is playing the piano' is true and 'Claire is buttering a toast' is true. Else (3) must be false. On the other hand a disjunction is true if and only if one of the two parts of the disjunction is true, or else the disjunction is false. That pre-theoretical (intuitive) considerations are absolutely necessary for an understanding of pathological expressions of natural language is supported first, by the necessity to rate proposed solutions of these problems and second, by the answer to the following question: With which logical laws should we describe circularity? The standard way to reason about circularity is classical logic. We will do the same here. In particular, we assume at this point that double negation of a proposition yields an affirmative proposition, i.e. we assume the availability of a classical negation.

We have to consider sentences and discourses which are circular in a particular sense. Most of our examples have the following property: under the assumption that the sentence (in the discourse/context) in question is true, the sentence must be false. On the other hand, under the assumption that the sentence (in the discourse/context) in question is false, then we have to conclude that the sentence has to be true. The paradoxical behavior can be alternatively described as follows: the pathological sentence in question is true if and only if the sentence is false. Hence, we have a classical contradiction. Some examples of paradoxical sentences and classical paradoxes are the following statements:[4]

(4)(a) This sentence is false. (Liar Sentence)
(b) This sentence is not true. (Strengthened Liar)
(c) The proposition expressed by this sentence is not true.[5]

Notice: In order to deduce a contradiction from (4)(b), it is not necessary to assume that (4)(a) and (4)(b) are equivalent.[6] For a better understanding of the examples we check the reasoning which yields the paradoxical conclusion (4)(b): Assume (4)(b) is true, which is equivalent to 'This sentence is not true' is true. Therefore, it holds: 'This sentence is not true'. The expression 'This sentence is not true' claims that (4)(b) is not true, which is inconsistent with our assumption. Therefore we have to assume that (4)(b) cannot be true. So we have to conclude that 'This sentence is not true' is not true. This is exactly what (4)(b) expresses. Hence, (4)(b) must obviously be true, which contradicts our assumption that (4)(b) is not true. We have to conclude: (4)(b) must be paradoxical, because (4)(b) is true if and only if (4)(b) is not true. The reasoning for (4)(a) and (4)(c) is similar to the above argument.

A widely accepted explanation concerning paradoxical sentences is the claim that these sentences are self-referential. As a consequence of this claim we get that avoiding self-reference of sentences avoids paradoxes in natural languages. As a counterexample to this thesis we consider example (5): The context (5) is a discourse in which no sentence is a Liar-like sentence, in particular no sentence of (5) itself is self-reflexive, but nevertheless we are faced with a paradoxical behavior of discourse (5):

(5) First sentence: The second sentence is true.
Second sentence: The third sentence is true.
⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮
n^{th} sentence: The $n + 1^{\text{st}}$ sentence is true.
$n + 1^{\text{st}}$ sentence: The first sentence is false.

[4] Different versions of the classical Liar sentence can be found in classical philosophical works. For example, in Buridan's 'Sophismata', chap. 8, 11th sophism: "What I am saying is false." It is obvious that Buridan's example is a version of the classical Liar sentence.

[5] We will assume that Liar-like sentences do express an ordinary proposition. If we deny this assumption, then the problems which arise with circular expressions can be solved very easily: one can develop a theory where only propositions can be true or false (not sentences). Without trying to discuss the problems of such a 'solution' here, it should be noted that this move is closely related to the old question which entity is the bearer of truth: a sentence, a proposition, or something else?

[6] With equivalence we mean that the predication 'is not true' is equivalent to the predication 'is false'.

Example (5) is usually called the Liar circle. Notice that no single sentence itself is self-referential,[7] but nevertheless we have a certain kind of circularity in (5): every sentence refers to the next one, the last to the first one. We get a vicious circle, and no definite truth value for the first sentence can be consistently assigned.

A natural question, concerning the properties of paradoxical sentences in natural languages arises: what happens, if we connect Liar-like sentences with other non-paradoxical sentences by standard logical connectives. In (6), some examples are formulated:[8]

(6)(a) John is eating his Hamburger and this sentence is not true.
(b) John is eating his Hamburger or this sentence is not true.
(c) If John is eating his Hamburger, then this sentence is not true.
(d) If this sentence is not true, then John is eating his Hamburger.

Example (6)(a) seems to be a sentence which is not in every situation (in other words in every possible world or in every possible situation) paradoxical: if it is the case that John does not eat a Hamburger, then the sentence seems only to be false, not paradoxical. This is important, because this stresses the fact that pathological behavior of sentences is not something that is a purely intrinsic feature of a language. Often, the paradoxical behavior depends on the context, in other words the way the world is. One must mention Saul Kripke as the first one who stressed this fact.[9] In our example, (6)(a) becomes pathological, if it is the case that John is eating a Hamburger.

In example (6)(b), we have the following analysis: if John is eating a Hamburger, then the sentence should be true, otherwise we get a paradoxical expression. Assume (6)(b) is true. Then one of the disjuncts must be true. The first one cannot be true because we assumed that the first disjunct is false, therefore the second disjunct must be true. Then it is true, what the second disjunct claims, namely that (6)(b) is not true. Now, if we assume that (6)(b) is not true, both disjuncts must be false. The first one is false by assumption, but the second forces us to conclude that (6)(b) is again true. We have a circle, or: (6)(b) is true if and only if (6)(b) is false.

(6)(c) is partially paradoxical: if we assume that the antecedent is true, then the whole implication is only true if the consequence is true. If we assume that the consequence is true, then it must be the case that the sentence must not be true which contradicts our assumption that the implication is true. So we have to conclude that, if the implication is not true, then the whole sentence does

[7] There is a second argument against the thesis that reflexivity causes paradoxes. In natural languages, there are various examples of sentences which are self-reflexive, but have no pathological properties. Consider the example: "US Air flight 205 will depart from gate H12 at six o'clock; this announcement will not be repeated." (Cf. [BarMo96]; p.55).

[8] In the following examples, it is assumed that the noun phrase 'this sentence' refers to the whole sentence. Without this assumption, we are able to formulate Liar sentences, too, but we have sometimes a different analysis of the sentences in question.

[9] Cf. [Kr75].

not have to be true; but this is exactly what the consequence claims. Therefore, the consequence is true: we have a contradiction. The situation changes if we assume that the antecedent is not true. Then, the implication is true and the consequent has no effect to the truth value of the implication. In this case, the truth conditions depend on empirical facts and the Liar sentence has no impact to the whole sentence.

(6)(d) is an interesting case. If we assume that the complex sentence is not true, then obviously the antecedent must be true and the consequence must be false which implies that the sentence is not true. If John is not eating his Hamburger, then it seems to be the case that the sentence has a classical truth value, namely (6)(d) is plainly false. Clearly, if John is eating a Hamburger, then we have a contradiction. Now, we assume that the whole sentence is true. Then, we automatically have a false antecedent and therefore the implication holds trivially: no paradox arises. Particularly in this case, it seems to be true that the interpretation of the whole sentence does not depend on empirical facts.

At this point, a remark with respect to ambiguities of sentences including Liar-like parts is useful. Clearly, the examples in (6) have different properties, if one restricts the scope of the noun phrase 'this sentence' to a relevant part of the complex sentence. The reader may figure out which properties would change if we restrict the scope of the demonstrative pronoun in an appropriate way.

Our next example was first formulated by Anil Gupta.[10] This example presents a discourse that cannot be modeled in Kripke's account of truth.[11] Whereas in Kripke's account (7) comes out to be paradoxical, this is not true for ordinary reasoning about discourse (7). In fact, (7) is not even pathological in any sense. Assume that A and B are two card game players who utter the assertions in (7) about a particular card game. We assume that two additional card game players C and D are playing and furthermore that D has the ace of clubs.

(7) A claims (a1)-(a3):
(a1) C has the ace of clubs.
(a2) All claims of B are true.
(a3) At least one of the claims made by B is false.
B claims (b1) and (b2):
(b1) D has the ace of clubs.
(b2) At most one of the claims made by A is true.

Example (7) is, according to Gupta, a non-pathological discourse, because we can reason in the following way: The claims made by A are inconsistent because statements (a2) and (a3) contradict each other, and therefore at least

[10] Cf. [Gu82].

[11] Compare Chapter 4 for Kripke's theory and Chapter 6 for problems of Kripke's partially defined truth predicates.

one of A's claims - (a2) or (a3) - must be false. We know that (a1) is false, because D has the ace of clubs (this is our assumption), and therefore (b1) is true and (b2) must be true, too, otherwise we were inconsistent. The uniquely remaining possibility is that (a2) is true. Conclusion: ordinary reasoning yields the result that there is no paradox at all, although the discourse is in a certain sense circular (and in the sense of Kripke's account ungrounded). In Kripke's fixed point approach, the described reasoning would be invalid, because we were not able to calculate that at least one of the assertions (a2) and (a3) is false (because they are inconsistent). The reason is that in Kripke's fixed point approach we are working in a partial logic in which it is possible that both statements have no definite truth value. Then, no inconsistency would arise, although (a)(2) and (a)(3) are contradictory. We will consider this example again in the context of the discussion of Kripke's approach towards a theory of truth.

In an short overview with respect to paradoxical sentences, the historically most famous example must be mentioned, even if it is not a very good example for the pathological behavior of pathological expressions of natural language. Consider example (8):

(8) A Cretan says: "All Cretans are liars."

The question is: Is the statement of the Cretan true or not? It is important to notice that (8) becomes only paradoxical in the case we assume additional information (which is far from being intuitive). First, we need to assume that all other Cretans (except the one who is uttering (8)) are liars. If we do not assume this, then the statement in (8) is plainly false, but not paradoxical. Second, we need to assume that a liar is a person whose statements are always untrue. Third, we assume that all utterances of the Cretan in (8) made before the utterance in (8) were untrue. Then, we get a paradox. Assume the utterance in (8) is true, then all Cretans have lied their whole life, in particular the Cretan in (8) has lied his whole life. But then, (8) cannot be true. Therefore, the utterance in (8) must be false. Then, there is at least one Cretan who said a true sentence in his life. According to our first assumption, this must be the Cretan in (8). But he was a liar his whole life according to our third assumption. So, his actual utterance in (8) must be true. Then, the circle starts from the very beginning. This example makes clear that background information (context) is sometimes an important ingredient in order to evaluate pathological sentences.

A famous example of sentences referring to each other in a paradoxical way (at least under certain circumstances) is the following dispute between Nixon and Jones (first mentioned in [Kr75]):

(9) Jones says: "Most of Nixon's assertions about Watergate are false."
Nixon says: "Everything Jones says about Watergate is true."

At first sight, we can see nothing paradoxical in this example. Furthermore, in most situations, when Jones and Nixon would utter these assertions, there is rarely a chance to derive an inconsistency. If we assume that the above statement spoken by Jones is his only one with respect to Watergate, whereas Nixon says the above claim and n further statements, such that $n/2$ of his further statements are false and $n/2$ are true, then we get a paradoxical situation.[12] The important insight of Kripke was that certain paradoxical sentences are not independent of facts in the world. Situations or the way the world is can transform an ordinary sentence with a definite truth value into a pathological expression of natural language.

In general, we would like to have a truth theory which supports a principle like Aristotle's principle of the excluded middle. For every sentence ϕ of natural language we would like to state a tautology like: 'Either ϕ is true or ϕ is not true and nothing else'.[13] What we intuitively do not want is a truth theory which lacks such an intuitive principle. This is relatively clear for ordinary sentences. What is the truth value for paradoxical sentences? Are these sentences plainly false, neither true nor false, or both true and false? The fundamental question is what the intuitions are with respect to sentences like (10)(a)-(10)(c) (in which λ represents the Liar sentence):

(10)(a) This sentence is true or this sentence is not true.
(b) This sentence is true or it is not the case that this sentence is true.
(c) λ is true or λ is not true.

Using ordinary reasoning, we can argue in the following way: (10)(a) is true if one of the disjuncts is true. Clearly, if (10)(a) is true, then the first disjunct has to be true and we are consistent. The whole sentence cannot be false, because then the second disjunct would be true and therefore the whole sentence would be true. Notice that (10)(a) is equivalent to (10)(b): This is caused by our 'implicit' assumption of the excluded middle, and the restriction to two truth values.

What are the differences between (10)(a)/(b) on the one hand and (10)(c) on the other hand? Are there any differences? Relative to our pre-theoretical intuitions this question can become crucial for the development of a theory of truth. Consider (11):

(11) 'This sentence is not true' is not true.

[12] Assume that Jones assertion is true. Then, Nixon's statement must be false, and therefore Jones utterance cannot be true. Now, assume Jones utterance is false, then Nixon's statement in (9) must be true, and therefore Jones statement is true. The circle starts again from the beginning.

[13] Intuitively, the principle of bivalence seems to be a promising principle, because it is implicitly used by everybody all the time. The question is whether a theory of truth allows such a principle.

The quoted sentence in (11) is the Liar sentence and because of its property of being paradoxical the statement that the Liar sentence cannot be true has to be true. Is this a sufficient reason to say (11) is true? If we make such a judgment, then we are able to conclude that (10)(c) is true: the second disjunct would be a tautology and therefore the complex sentence must be true. As a consequence of our considerations we would get the equivalence between (10)(a) and (10)(c). This kind of evaluation of problematic sentences seems to be correct for many researchers,[14] because we preserve classical tautologies and we are able to work with Liar sentences without collapsing our system in the manner that whenever a Liar sentence is contained in a complex sentence, then the whole sentence becomes paradoxical. The last conclusion would be quite implausible, because one would be led to a standpoint not far away from a position where only languages which are not strong enough to represent circular sentences (or discourses) can be represented. From that perspective it seems to be reasonable to try to create a theory of truth validating a principle like (12):

(12) $\forall\phi \in L$: (ϕ is true or ϕ is not true)

As a side remark, it is to mention that (12) is not valid in Kripke's fixed point approach if L is a sufficiently strong language. Expressions like (10)(c) or a sentence of the form $\forall\phi : (\mathbf{T}(\phi) \vee \neg\mathbf{T}(\phi))$ are necessarily pathological in his system. The reason is the evaluation of paradoxical expressions.

There are not only paradoxical sentences in natural language. There are additionally sentences that are pathological in their behavior but nevertheless not Liar-like in character. The classical example of a non-paradoxical sentence which is nevertheless pathological is the so-called 'Truth-teller' or 'same-sayer' sentence. Here are some versions of the 'Truth-teller' sentence:[15]

(13)(a) This sentence is true.
(b) The proposition expressed by this sentence is true.
(c) What I am now saying is true.

If we assume that examples (13)(a)-(c) are true, then the sentences remain true. If we assume on the other hand that they are not true, then they remain not true. Nothing changes in the evaluation process or to put it more intuitively: there is no flipping behavior of (13)(a)-(c). Notice that although we are not involved in inconsistencies, it is not possible to assign a definite truth value to one of the expressions in (13). (At least this is not possible in an intuitive convincing way.)

[14] Intuitively, it is not the case that (9) is equivalent to the statement 'The Liar sentence is true' (which would be clearly false), because this sentence would be fully formulated with 'This sentence is not true' is true.

[15] Again we assume (as in the Liar case) that the noun phrase *this sentence* in (13)(a) and (13)(b) refers to the whole sentence itself.

Sentences which claim their own truthfulness can be embedded into complex sentences (as well as in the Liar case) with the usual propositional connectives. One important and very interesting example[16] is the following one:

(14) If this sentence is true, then John is playing the piano.

If we assume that (14) is true, then the antecedent has to be true and the whole sentence is true if and only if the consequent is true. Therefore it seems to be the case that we are able to derive that John is playing the piano, using conditional proof techniques. In the other case, where we assume that the whole sentence is false, we have an interesting behavior: it cannot be possible that this is the case, because then the antecedent is false and therefore the implication is true. This has a strange consequence: If we consider sentences like (14), it seems that we are able to prove that something is the case in reality (in our case that John is playing the piano), although there is no intuitive justification for such a result based on the facts that do hold in the world. This example has its defective properties not in being contradictory, but in the possibility to say something about the world where there is no connection between the world and our interpretation supporting our conclusion.

We end this section by mentioning a terrifying story about a logician captured by cannibals:

> (15) The cannibals hoped to have a good meal for the evening. Although they ate human beings, they were very truth-loving people and they were not cruel: they permitted the logician to choose his manner of death, by decapitating him, if he says something true and by torture him to death, if the last statement of his life is false. The logician thought a few seconds about his alternatives, smiled and said: "You will torture me to death." The cannibals were forced to set him free.[17]

The dilemma of the cannibals (and the fortune of the logician) is the following: if the statement of the logician was true, then the cannibals had to decapitate him, but then the claim was clearly false, and the cannibals should have tortured him to death: contradiction. If the statement of the logician was false, then the cannibals would have to torture him to death, but then the statement would be true and they could not torture him to death. Therefore, their only possibility was to set him free.[18] Now we will

[16] We will not mention and describe other embeddings of Truth-teller sentences in complex sentences, because the treatment of these is quite similar to the Liar-like examples.

[17] Obviously, the puzzle would work with a statement by the logician like the following one as well: "The sentence I say now is false." Then we have a pathological statement quite similar to the examples above. In our example, the nice property is that we have no circularity or self-referentiality. The only problem is that the truth value of the sentence is not fixed and cannot be fixed using the possibilities provided by the cannibals.

[18] Again this is an example of the tremendous importance of logic in daily life.

assume that the logician does not answer "You will torture me to death" but "You will decapitate me". The question is, what will the cannibals do? If they torture the logician to death, they are consistent in their behavior, if they decapitate him, they are consistent, too. Is this a case where we have to assume that the answer in question given by the logician is both true and false? Notice: the cannibals are not able to decide which kind of execution is the right one. Each choice is necessarily an arbitrary stipulation. What would be a rational choice in this case? An answer depends on personal attitudes. In a certain sense, the cannibals should let the logician go because there is no rational decision as to what kind of execution is the right one. On the other hand, every execution is consistent, so the cannibals can choose. I do not dare to think what the cannibals will probably do.[19]

In the next section, we will consider circularity in the context of knowledge representation.

2.3 Circularity and Knowledge

In the following considerations, we will examine whether circularity can be detected in knowledge reports, human reasoning, and knowledge representation. To do this we will state certain puzzles, which have been known for a long time. The first one has many different forms and is usually called the Conway-paradox. The most famous formulation is probably the muddy children problem. Because we want to model the muddy children problem later, we present this example quite explicitly:[20]

> (16) Suppose that n children are playing in the garden. Their mother told them that they will be subject to severe punishment when coming home dirty. Now, every child tries to make the other child dirty at the forehead, because the dirty child cannot see that it is dirty and so the mother will punish him/her. After a certain time playing in the garden, k children are dirty on their forehead. The father comes along and makes the following statement: "At least one of you is dirty on the forehead. Can anyone definitely say that he/she is dirty?" None of the children can say anything definitely to the father, because they cannot see their foreheads. The father asks the question again: "Can anyone definitely say that he/she is dirty on the forehead?" After repeating his question k times, the k dirty children say together: "Yes, I am dirty."[21]

[19] It is clear that the first answer of the logician corresponds to the Liar sentence, whereas the second utterance corresponds to the Truth-teller sentence.

[20] The puzzle as stated here can be found in [Bar81]. This paper made the puzzle famous.

[21] We assume that the children are rational.

How is it possible that although the father does not give additional information after the first question, but simply repeats his question again, the children can reason correctly? What is going on here? It is helpful to consider the reasoning of a dirty child. For reasons of simplicity, we assume that there are only two dirty children. After the first question of the father, the child reasons that there is another child who is dirty (because he/she can see the other dirty child), but his/her own status remains unclear. So, the child is not able to give a reasonable statement about his/her own situation. After the second question of the father, the child reasons as follows: He/she sees only one child who is dirty (assumption). Then after the first question of the father, this particular child had answered if he/she had been the only one. So, there must be another dirty child. And because he/she cannot see any other dirty child, the child himself/herself must be the one in question.

We should emphasize that there is a certain shift in the type of knowledge of the child: at the beginning, he/she knows that at least one child is dirty. After the first question of the father and the fact that the other dirty child does not answer to the question of the father, he/she knows that the other dirty child knows that someone else is dirty. At this point, the knowledge that there is not only one dirty child becomes public knowledge. And a way to say this is: every child knows that every other child knows that there are at least two dirty children. This is called public knowledge. We will see later that an analysis of public knowledge requires a circular treatment.

We will consider another example. The following puzzle has a very similar structure as (16), but stresses the difference between private and public knowledge more clearly. We quote this puzzle from [Bar90], p.201:

> (17) *Suppose you have two poker players, Claire and Max, and each is dealt some cards. Suppose, in particular, that each of them gets an ace. Thus, each of them knows that the following is a fact:*
>
> s: *Either Claire or Max has an ace.*
>
> *Now suppose Dana were to come along and ask them both whether they knew whether the other one has an ace. They would answer 'no', of course. And if Dana asked again (and again...), they would still answer "no".*
>
> *But now suppose Dana said to them, "Look, at least one of you has an ace. Now do you know whether the other has an ace?" They would again both answer "no". But now something happens. Upon hearing Max answer "no" Claire would reason as follows: "If Max does not know I have an ace, having heard that one of us does, then it can only be because he has an ace." Max would reason in the same way. So, they both figure out that the other has an ace.*

The important property of the above puzzle is, that although Max and Claire do not know more than before, after having heard from Dana that at least one of them has an ace, they perform different conclusions in reasoning about the given situation. In the first case, the knowledge that at least one of them has an ace is private knowledge, whereas in the second case it is common knowledge. This means that each of them knows that the other knows the same as her- or himself. This makes the whole difference, and has enormous consequences concerning their ability to perform conclusions. Our aim will be to find ways to model these phenomena. It will turn out that what is going on in the process of reasoning is a circular process.

We mention a last example. The following puzzle seems to be folklore for people who like puzzles.

> (18) Assume that three logicians are captured by cannibals. The cannibals do not want to be cruel. So, the logicians are allowed to solve a puzzle in order to free themselves. The cannibals posit five stakes in a linear order behind each other. The logicians are tied on the first three stakes, such that the third logicians can see the first and the second stake, the second logician can see the first stake and the first logician cannot see any other stake. Three of the five stakes are white, two are black. The task is that one of the logicians must tell the cannibals the color of his/her stake. Then, all three logicians will be free. Is the guess of the logician false, then all three have to die. The logicians figured out the strategy to solve this problem. The question is: How?

For a solution one has only to check the possibilities: if the first and the second stake is black, then the third logician knows that his/her stake is white. If the first stake is black and the second is white, then the second logician knows that his/her stake is white, because the third logician does not say anything. In all other cases, the first logician knows that he/she is on a white stake. Therefore, there is a unique solution to this problem. The question is how this reasoning can be modeled.

An analysis of common ground can be formulated as follows.[22] Assume a certain situation s is given. We assume further that s supports a certain proposition p about situation s. Assume additionally that a discourse community consisting of n people know p, but additionally, they know that everybody in the discourse community knows p, too. As we saw in the example of the muddy children puzzle, sometimes it is necessary for a member k of the discourse community to know that member k' knows that k knows that p. Or that k

[22] There is not only one possibility to analyze common ground. We adopt here an interpretation which goes back to an analysis of David Lewis (Cf. [Le69]).

knows that k' knows that k'' knows that p. The possibility to iterate these that-sentences arbitrarily often is one essential feature of common ground. What is a reasonable analysis of situation s? One modeling of s can be given as follows: We have to sum up the propositions (facts) that hold in s. A necessary but not sufficient condition is that s is a situation that includes the proposition that every member of the discourse community knows p. Additionally, and that is where the circularity comes in in this type of modeling, s is a situation in which every member knows the situation s itself. If we want to model a situation as a set of propositions,[23] there is no set in standard ZFC set theory that models this situation. The problem is that our situation s includes itself in one or the other way. Or in other words: s cannot be specified without an explicit reference to s itself. This is surprising, because common ground and common knowledge seem to be quite ordinary in our life. Nothing seems to make common ground phenomena paradoxical in the sense the Liar sentence is paradoxical.[24] The moral of this example can be formulated as follows. Circularity is a phenomenon that happens to occur quite often in different aspects of life, without disturbing our reasoning or making our belief contexts inconsistent.

In the next section, we will consider occurrences of circularity in other fields like in linguistics, mathematics, logic, and computer science.

2.4 Circularity in other Fields

Circular phenomena are present in various academic disciplines like in the foundations of mathematics, in mathematics in general, in computer sciences, and in linguistics. We want to focus on some important occurrences of circularity that are prominent on the one hand because of their conceptual consequences, and on the other hand because of their historical impact on the field. Again, we do not claim to give a complete list of phenomena that can be found in these disciplines. We only give an overview. We begin with some linguistic examples.

2.4.1 Circularity in Linguistics

The semantical paradoxes in Section 2.2 can count as linguistic examples. The problem of semantical paradoxes challenges formal semantics to integrate an appropriate truth predicate. For linguistic applications, it is quite important to have a model that can represent the truth predicate.

Intuitively, grammatical sentences of natural language seem to be non-circular concerning their syntactical structure. In general, circular anaphoric reference is not possible. The following examples make this claim clear.[25]

[23] We can assume that propositions can be modeled as sets.

[24] Clearly, we adopted here a special analysis of common ground. There are other possibilities like an inductive account. In an inductive account, no circularity arises, but we have to deal with infinite objects. In Part IV of this work, we will see that under certain circumstances circular representations can be unfolded using infinite objects.

[25] For the following examples compare [Ha91].

(19) Rebecca gives Carl the book.
(20) Rebecca gives her husband the book.
(21) His wife gives Carl the book.
(22) His wife gives her husband the book.

Whereas in (19)-(21) a reading is possible where Rebecca and Carl (or the pronoun plus noun) are spouses of each other, this reading is impossible in (22). In other words: in (20) and (21) the pronouns can have their antecedent in the sentence. This is impossible in (22). It seems to be the case that in ordinary contexts circular reference of the pronouns is impossible:

(23) *His$_i$ wife$_j$ gives her$_j$ husband$_i$ the book.

The above examples seem to justify that reference of pronouns is well-founded. Unfortunately, this is misleading under a certain interpretation and certain assumptions. There are grammatical sentences where anaphoric reflexivity is possible under an appropriate interpretation. The famous Bach-Peter's sentences are examples. Consider the following sentence (24).

> (24) The secretary who knew him shouted to the farmer who did not recognize her.

Let us consider the reading with the circular reference of the pronouns, i.e. the reading where every pronoun has its antecedent in the sentence. In this reading, the identification of the noun phrase *the secretary* is dependent on the discourse referent of the pronoun *him.* This pronoun is dependent on the phrase *the worker who did not recognize her.* In order to identify *the worker*, we have to refer to the discourse referent of *her*, but the reference of this pronoun is essentially dependent on the interpretation of *the secretary.* And at this point the circle starts again.[26]

Notice: although (24) is obviously grammatical, the reference-structure of the discourse referents can be interpreted as a circular phenomenon. Although we will not analyze Bach-Peter's sentence in more detail, we should mention that in the standard syntax theory, government and binding theory, circular references are blocked by a principle. It should be mentioned that there are explanations of Bach-Peter's sentences in GB that do not mimic the circular character of the reference structure.[27]

2.4.2 Circularity in Mathematics

Our next examples for circularity come from the wide field of mathematics. An obvious and apparent usage of circular sentences to prove important results

[26] It should be mentioned that we need certain assumptions in order to perform this reasoning. First, it is necessary that the semantics is interpreted compositionally. Second, a certain logical syntax is needed to get this interpretation. In contrast, a straightforward representation of Bach-Peter's sentences in first-order predicate logic is not circular.

[27] Compare for example [Hi80].

in the foundations of mathematics can be found in the famous incompleteness theorem of Kurt Gödel. Another famous theorem where circularity plays an important role is the theorem of the undefinability of truth in sufficiently strong enough languages by Alfred Tarski.[28] Examining these results, the question arises where the circularity comes in. Gödel's incompleteness theorem claims that there is a sentence of arithmetic that is true (in the standard model) but that is not provable in Peano-Arithmetic. Additionally, its negation is not provable in Peano-Arithmetic as well. In order to prove this theorem, Gödel explicitly gave an example of a sentence that has precisely these properties. Intuitively and strongly simplified, Gödel's sentence claims that the sentence itself is not provable in arithmetic. The following dilemma occurs on an intuitive level: if the sentence is provable, then the sentence must be false in the model which is not possible. If the sentence is not provable, then the sentence is true (because it claims its own unprovability). Hence, the system must be incomplete.

The importance of Gödel's theorem cannot be underestimated. The consequences for the foundations of mathematics were enormous. For example, Gödel's theorem made clear that Hilbert's program of 'reliable mathematics' cannot be successfully accomplished.[29] Additionally, in a certain interpretation of the incompleteness theorem not only a principal borderline of the concept of provability can be established but a principal borderline of human reasoning as well. Nobody is able to go behind certain axiomatic theories because of principal reasons. Finally, complexity theory is not conceivable without this result.

Tarski's theorem of the undefinability of truth in sufficiently strong languages uses a self-referential sentence, too. Tarski constructed a sentence that claims of itself to be an element of a theory (self-reference lemma). This sentence can be used to prove that truth is not definable in languages that contain at least arithmetic (or more generally an appropriate elementary coding schema which is quite similar to arithmetic). Tarski's theorem was very important for the development of type theory and for the development of a theory that specifies what it means for a sentence to be true in a model. Additionally, because of Tarski's theorem the examination of higher-order logics became more important. The reason for this is that one can show that truth of first-order theories can be defined in second-order logic.

The most famous paradox in the foundations of mathematics was found by Bertrand Russell. He recognized that naive set theory with unrestricted comprehension axiom is inconsistent. There are nice and intuitive representations of the Russell paradox in natural language. One version can be found in [Sa95]. In this version, the Russell paradox has the following form: Assume that there is a barber in a small village in Italy who shaves exactly the persons in the village that do not shave themselves. This works for all people except the barber

[28] Compare [Go31] and [Ta56].

[29] Some researchers claim that this thesis is too strong. We do not want to discuss the influence of the incompleteness theorem on Hilbert's program in detail here.

himself: who shaves the barber? If he shaves himself, then the above constraint implies that the barber is a person who does not shave himself. On the other hand, if the barber does not shave himself, then he would need to be shaven by the barber (because the barber shaves exactly the persons in the village who do not shave themselves). We are in a dilemma. An intuitive explanation of the above situation can be given by the claim that a barber characterized by the above constraint does not exist in the village.

In set theory, we have the following version of our little story of the barber in an Italian village. Consider the set X that is characterized by the following definition:

$$X = \{x \mid x \notin x\}$$

The problem of the last collection becomes obvious, if one asks whether X itself is an element of X or not. Assume it holds $X \in X$. Then the following reasoning is valid:

$$\begin{aligned} X \in X &\Rightarrow X \in \{x \mid x \notin x\} \\ &\Rightarrow X \notin X \\ &\Rightarrow X \notin \{x \mid x \notin x\} \\ &\Rightarrow X \in X \end{aligned}$$

This implies that it holds $X \in X \Leftrightarrow X \notin X$, which is a contradiction. The standard way out of these kinds of paradoxes in set theory (the Burali-Forti paradox in the theory of ordinals is of a very similar type) is to prevent definitions like $X = \{x \mid x \notin x\}$, because these definitions do not define a set of the universe. The formal axiomatization of ZFC drops the unrestricted comprehension axiom and adds the separation axiom. This modification guarantees that these collections do not form sets, but are objects of a different type, namely proper classes. And, in fact, there are classes that contain themselves: for example, the collection X of all objects that are not coins must include itself as an object, because the collection X is obviously not a coin. Classes can intuitively be interpreted as sets that are too big in order to be 'real' sets of the universe. Therefore, mathematicians introduced a new ontological category (type) in their framework, namely classes. Notice that the mentioned standard solution in ZFC set theory (namely to drop the comprehension axiom and to add the separation axiom) uses the strategy to prevent that certain collections are interpreted as sets. This is precisely the solution for the barber problem. We deny the existence of such an object. There are other possibilities to give reasonable solutions to the Russell Paradox, solutions that change the concept of a set in other respects. We will not consider these possibilities here and refer the reader to the two classical works [Ac78] or [RuWh27].

We want to end our discussion of examples of circularity in mathematics with occurrences of circularity in some very obvious and basic mathematical formulations. First, we make the following observation: On the one hand, circularity, or more precisely self-referential sentences, were used to prove some of the most important results in the foundations of mathematics. On the other hand, circular concepts can be used to introduce new mathematical objects. We will see a set theoretical example later: the introduction of circular sets (hypersets) can be based on solutions of set theoretical equations (or more generally systems of equations). Very similar constructions occur in very basic mathematical theories, too. Let us consider a trivial polynomial equation.

$$px = q$$

We assume that p and q are integers. Now consider the pair $\langle a, b\rangle$ that corresponds to the equation $ax = b$. We define equivalence classes on these pairs according to:

$$[\langle a, b\rangle] \equiv [\langle c, d\rangle] \ \Leftrightarrow \ ad = bc$$

If we define 1 as the equivalence class $[\langle 1, 1\rangle]$ and 0 as the equivalence class $[\langle 0, 1\rangle]$, then one can check quite easily that this construction satisfies the axioms of a field. Now we can calculate for every $px = q$ that $x = [\langle q, p\rangle]$ where $[\langle q, p\rangle]$ is the equivalence class of $\langle q, p\rangle$ as defined above. In other words: We have introduced the rational numbers via solving algebraic equations.

The important difference between the introduction of integers and the introduction of hypersets (compare Chapter 11) is the fact that in the above algebraic example no real circularity occurs. The indeterminate x does not occur on both sides of the equation, whereas in the set theoretic extension of ZFC we will see later exactly that will happen.

Even in ordinary mathematics we can find this kind of circularity, for example, in the theory of ordinary differential equations. The following equation has a unique solution.

$$\frac{d^2f}{dx^2} + a\frac{df}{dx} = f(x)$$

In a certain sense, this equation is circular, because on both sides of the equation we have occurrences of the function symbol f (even though on the left side the occurrences of f are occurrences of derivatives of f). Standard techniques of the theory of ordinary differential equations guarantee that the solution can be effectively calculated. In a certain sense, our example is a 'good' form of circularity, because we can solve the equation and we get a solution. Another 'good' form of equation is an equation of the form

$$f(x) = f(x + 1)$$

The solution class for this equation is the collection of all functions that are constant. Whereas we have a solution for this case, this does not hold for equations of the following form.

$$f(x) = f(x) + 1$$

There is plainly no function in mathematics that satisfies the above equation because a solution would contradict the definition of a function. What is the principal difference between the mentioned equations and their different behavior? The difference is that, in one case, the equation describes an 'object' that does not exist in mathematics, whereas in the other case there is a function (or number) that satisfies the equation. Like in the case of set theory and the definition of collections, we are confronted with the fact that certain 'definitions' make no sense, others do. Circularity is not necessarily a sufficient condition in order to create a 'bad' definition. Quite often, circularity can be used to strengthen a given theory and to introduce respectable objects in mathematical discourse.

2.4.3 Circularity in Logic

In a certain sense, most of the mentioned examples of circular phenomena are logical examples. In this subsection, we mention a paradox that occurs in predication theory. Historically, the paradox is due to Grelling.[30] It is structurally different from the examples above because we do not consider truth values of sentences, or self-referential sentences, but simply extensions of a predicate. Nevertheless, it is quite similar to the above examples as we will see in a moment.

Let us consider satisfaction relations concerning predicates. We call a word 'heterological' if it does not apply to itself. For example, 'long' is a heterological word, because it does not apply to itself, whereas 'short' is not heterological, because it applies to itself. (We assume an appropriate pre-understanding that words that have equal or less than five letters are 'short'). The problem is this: is 'heterological' a heterological expression or not? If it is heterological, then it does not apply to itself, which means that 'heterological' cannot be heterological in accordance with the definition of heterological expressions. Therefore, we have to assume that 'heterological' is not heterological, that is: it does not apply to itself. But this is the definition of heterological, therefore 'heterological' is heterological. We have to conclude: 'heterological' is heterological if and only if 'heterological' is not heterological. Again, we are facing a paradox.

One strategy to prevent this paradox is to apply our set theoretic account. Predicates like 'heterological' simply do not exist in our universe of discourse. The similarity between set theory and predication theory is not so surprising, because the extension of a predicate is usually interpreted as a set of those objects that satisfies the predicate. (Exactly this is the point where we have the similarity to the comprehension axiom in naive set theory). A way out would be to introduce predicates of a higher type, very similar to set theory where classes are these collections.

[30] The example is taken from [Vi89], p.621. The paradox was found by the mathematicians Grelling and Nelson in 1908 (compare [GrNe08]).

2.4.4 Circularity in Computer Science

We finish this section with some remarks concerning circularity in computer science. We choose a very informal form for the presentation of this introductory part. Unfortunately, it is not easy to present some interesting features about circularity in computer sciences without becoming quite formal and technical. Therefore, we refer the reader to [BarMo96, Rut95, Rut96, MiTo91] for a more explicit and larger overview. These works are also the place to find further papers about this topic.

Our first example is the concept of streams. If a set A is given, a stream $s = \langle a, s' \rangle$ is a pair such that $a \in A$ and s' is another stream. If one wants to speak about properties of streams, it is useful to be able to extract from a given stream $s = \langle a, s' \rangle$ the first element 1^{st} and the second element 2^{nd}. Let 1^{st} and 2^{nd} be functions extracting from a given stream s the first and the second element, respectively. This corresponds to projection functions in recursion theory. Now the following problem arises: if s is of the form $s = \langle a, s \rangle$ (or more precisely, $s = \langle a, \ulcorner s \urcorner \rangle$ where $\ulcorner s \urcorner$ is a name for s), then it is not clear a priori that this stream has a set theoretical representation in ZFC. The problem is that in ZFC the foundation axiom does not allow collections that contain themselves as objects. At this point, the theory of hypersets can be used to give streams a general and natural set theoretical foundation. Additionally, if we represent the set of all streams over A by A^∞, then we have obviously: $A^\infty \subseteq A \times A^\infty$. At the same time, $A \times A^\infty \subseteq A^\infty$ should also hold, because a pair $\langle a, s \rangle$ where $a \in A$ and $s \in A^\infty$ seems to be a stream in its own right according to our definition of streams. So the collection of all streams should satisfy the equation $A \times A^\infty = A^\infty$.[31] This set A^∞ does exist in ZFC, namely the collection of all infinite streams. This collection is in fact a set of the universe.

Let us assume for a moment that we already have a set theory in which such non-well-founded collections like $s = \langle a, s \rangle$do exist, referring the reader to later chapters. How can we define operations on single streams? For example, assume we want to define for a given stream

$$s = \langle n_1, \langle n_2, \langle n_3 ..., \rangle \rangle \rangle$$

the stream

$$s' = \langle 2n_1, \langle 2n_2, \langle 2n_3, ..., \rangle \rangle \rangle$$

where the n_is are taken from the positive integers. A natural idea to model this is to model the above relation between streams via a function map_f that is itself dependent on a function

$$f : \mathbb{N} \longrightarrow \mathbb{N} : n \longmapsto 2n$$

Then, map_f can be formulated as follows.

$$map_f : A^\infty \longrightarrow A^\infty : s \longmapsto \langle f(1^{st}(s)), map_f(2^{nd}(s)) \rangle$$

[31] An explicit discussion of these claims can be found in Chapter 13.

The reader can easily check that we can calculate:

$$map_f(\langle 1, \langle 2, \langle 3, ...\rangle\rangle\rangle) = \langle 2, \langle 4, \langle 6, ..., \rangle\rangle\rangle$$

That is the desired result. It is important to notice that the definition of map_f has a certain similarity to recursive definitions, although there is no real base case where the recursion can be founded. Later, we will examine various kinds of these, so called corecursive definitions. Additionally, we should mention an important feature of these kinds of definitions: we work with infinite objects. In ordinary recursion theory, although we can define recursively infinite sets of objects, each single object in the collection is usually considered as finite. In the above examples, we define explicitly infinite objects via corecursion, namely infinite sequences (streams). In later chapters, we shall consider the theory of corecursion that provides quite general techniques to define operations on streams (compare Part IV for further information).

A further example where circularity plays a role in computer sciences is the theory of labeled transition systems. Labeled transition systems can intuitively be understood as automata. We will consider the following definition specifying the concept of labeled transition system.

Definition 2.4.1 *Assume a set Act of actions is given. (The set Act is considered as 'atomic' in the sense that there is no further analysis of these actions.) A labeled transition system T is a pair $T = \langle S, \delta\rangle$, such that S is a set (representing states) and $\delta : S \longrightarrow \wp(Act \times S)$ is a transition function.*

Intuitively, if $\langle a, t\rangle \in \delta(s)$, then this can be interpreted as: given state s we can move from s via action a to state t. This can be represented as follows: $s \xrightarrow{a} t$. The transition function is not necessarily deterministic. We want to consider an example to make the situation easier to understand.[32] Assume we have an atomic set of actions given by $Act = \{beat, stop\}$, and we define T as the set $T = \langle\{heart, heart\ attack\}, \delta\rangle$ where δ is given by the following two equations

$$\delta(\{heart\}) = \{\langle beat, heart\rangle, \langle stop, heart\ attack\rangle\}$$

$$\delta(\{heartattack\}) = \emptyset$$

The described labeled transition system models the functioning of a heart. Our system *heart* can perform two transitions: either it can beat, after which it is still in the state *heart*, or the heart of the animal can stop beating. Then, the animal is in the state of a *heart attack*. (We assume here that there is no physician available to save the life of the animal.) If the animal is in the state *heart attack*, then there is no action to get out of this state. An important feature of our example is that we identify intuitively the system itself 'heart' with its behavior (properties) to be something that is beating. In the case that the system 'heart' is no longer beating, it is no longer a heart. Therefore, our

[32] The following example is taken from [BarMo96].

systems can be described by the following equation:

$$heart = beat.heart + stop.heart\ attack$$

Again we are faced with a problem to represent those examples in standard ZFC set theory, because there are no sets that satisfy the above equation. Again, hypersets can be used to model this state of affairs.

In the next section, we add some remarks concerning circularity occurring in philosophical contexts. The rejection of an argument in philosophical discourse because of a vicious circle in the argument is quite common.

2.5 Circularity in Philosophical Explanations

2.5.1 Scepticism

In this section, we examine some features of circularity in philosophical debates. It seems to be the case that circularity occurs relatively often in philosophical contexts, especially in order to show that someone's argument is in one or the other way 'begging the question'. Good examples can be found in the history of analytic philosophy. For example, consider the famous argument of G.E. Moore to show that there is a physical reality and we are not brains in a vat (referring to Putnam's famous example to give an anti-skeptical argument that Moore did not know).[33] The standard criticism concerning Moore's argument of the existence of a physical reality is the claim that Moore is begging the question, because his argument contains a vicious circle. The argument can be formulated as follows:

> Premise 1: I know that I am eating a sandwich right now.
>
> Premise 2: If I know that I am eating a sandwich right now, then I know that I am not a brain in a vat in an otherwise empty world.
>
> Conclusion: I do know that I am not a brain in a vat in an otherwise empty world.

The standard criticism in philosophical debates concerning the above reasoning can be summarized as follows. The sceptic tries to question the reliability of perceptual knowledge in reminding us that it could be the case that we are brains in a vat. Moore is rejecting scepticism by an argument that crucially depends on Premise 1, namely the claim that I know that I am eating a sandwich right now. That is to claim that I know something about

[33] A good presentation of Moore's argument can be found in [Wa99], p.86. Usually, it is assumed that Warfield's presentation is a correct interpretation of the original argument of Moore.

my perceptual knowledge, in our example that I am eating a sandwich right now. So, to prove that perceptual knowledge is not misleading, we assume that perceptual knowledge is not misleading. This is a circular argument. Even though the formulation does not make this explicit (and the circularity is in a certain sense hidden), the circular aspect of Moore's argument seems to be quite obvious when we consider the basic claim behind the formulations of the premises and conclusion.

In an important respect, the same kind of circularity as in Moors's case arises in Putnam style arguments against scepticism. Without considering in depth the partially highly sophisticated and difficult debate concerning the scepticism/anti-scepticism arguments in recent philosophical discourses, we only examine the very basic form of Putnam's argument.[34]

Premise 1: I can truthfully think 'I am a brain in a vat'.

Premise 2: A brain in the vat cannot think 'I am a brain in a vat'.

Conclusion: I am not a brain in a vat.

Whereas premise 2 is essentially based on the assumption that semantic externalism is true, it is not sufficient for premise 1 to assume semantic externalism in order to be true. That holds, because semantic externalism makes a claim about references of (certain) notions, but does not provide us with a real basis to assume that we are really outside of a vat. To put it differently: Semantic externalism tells us that reference and meaning are not only in the brain but also at least partially related to entities outside of the brain. (To what extend the outside aspect comes into consideration varies, dependent on the particular philosophic position.) The important point is that in order to be able to think truthfully 'I am a brain in a vat', we have to presuppose that we are in fact not brains in a vat, because our notions refer to physical objects. In other words, the argument has a circular character and can be reduced to the reasoning: I am not a brain in a vat because I assume that I am not a brain in a vat.

We shall mention a further point in this context. It is obvious that in order to be able to state the above argument, the individual stating the argument must formulate it in a meta-language that is rich enough to speak about the situation of the brains in a vat as well as about the situation in the physical world. This is not possible for the brain in a vat by the assumption of semantic externalism. Nevertheless it is a necessary condition for the validity of the

[34] Compare for different explications of Putnam's claim [Pu81]. Other sources are [Wa99], p.78 or [Br92], pp.48. There are many different versions of this argument depending on the strength of the claim. Whereas Warfield tries to argue for a relatively weak claim, namely that only some a priori knowledge about my own reasoning is necessary, in Brueckner and Putnam's own papers it is quite obvious that the claims are much stronger. These authors seem to argue for the claim: 'We know that disquotational truth-conditions hold in our world'.

argument to be able to speak about both cases (and to compare them), i.e. the individual making the claim cannot be inside the vat. Therefore, we presuppose that we are not brains in a vat. Again, the circular character of our reasoning is quite obvious.

Historically, the philosopher who was considered as the first one who gave a strong anti-sceptical argument was René Descartes. Even his argument has a circular character as was shown in [BarMo96]:

> *"The reason is that Descartes' act of doubting itself requires thinking and Descartes was aware of this. Basically, Descartes' famous dictum is shorthand for something more like this: I am thinking this thought, and this I cannot doubt because my doubting requires my thought."*[35]

It seems quite obvious that every argument that is 'self-refuting' in character or a performative self-contradiction has a certain intrinsic circular character. In the last years, especially John Searle argued for circular features of large parts of intentionality. We cannot discuss this here and refer the reader to [Se83], but we want to point out the intuitive idea behind that, based on the analysis of intentions by [Gr66]. If a person A has intention I by uttering S, then this is usually explained by saying that A wants to produce a certain effect in the audience by making them aware of his/her intention I.[36] This is an essentially circular explanation of the philosophical term *intention*.

The rejection of philosophical arguments by philosophers because of a hidden circularity in the argumentation is very common in philosophy. In ancient times, one version of these rejections (occurring very often, especially in the writings of Aristotle) is the claim that a certain argumentation yields an infinite regress. Aristotle used infinite regresses quite often to show that certain possibilities cannot be true and must therefore be replaced by their contrary. Another good source for examples of this kind of arguments in philosophy can be found in the writings of the sceptic Sextus Empiricus.[37]

2.5.2 Explanations

Usually, an explanation is considered as bad if one uses a circular argument in the explanation. Most important theories seem to be well-founded in the sense that they assume a certain basis as given and are constructed on this basis.[38] As long as a theory is well-founded, we usually do not reject the theory because of methodological reasons. One can argue whether the assumed axioms are reliable or not, but one usually does not question the method.

Mathematics is essentially based on two formal concepts. First, set theory provides the ontological basis for mathematics. This basis enables the mathematician to speak about mathematical objects, such as functions, relations,

[35] Cf. [BarMo96], p.51.

[36] Clearly, this presentation is a littel bit simplified in order to stress the important point.

[37] Cf. [Se94].

[38] Quite often an inductive process is involved in the construction.

spaces, numbers, groups, or fields. Second, classical logic is used in order to prove theorems in mathematics. At this level, we are using a meta-language to be able to speak about objects of a particular theory such as group theory or the theory of algebraically closed fields. For example, properties of objects in mathematics are represented in this meta-language. There is an important second aspect of logic in mathematics, namely the aspect that we need to formulate set theory axiomatically. Whereas naive set theory provides us with no consistent framework to speak about sets, axiomatic set theory, for example ZFC, does. And ZFC is quite successful: most mathematical discourses can be formulated in ZFC. Now, ZFC is axiomatized in first-order logic and exactly here is it where a flavor of circularity of the foundations of mathematics comes into consideration. In order to state the proof-theoretical part or the semantical part of first-order logic, we need set theory. That is true because we need to code the objects in logic as 'mathematical objects', and this must be done using set theoretic tools. Set theory as formulated in ZFC is only available in an axiomatic version using essentially first-order logic. Logic cannot be used without set theory and set theory is essentially axiomatized using first-order logic. The problem can be described in other words as follows: logic and set theory are not established as mathematical theories independently. They are crucially based on each other in a reciprocal way.

The usual strategy to avoid this problem is to refer to the meta-language character of set theory when first-order logic is established, and on the other hand the meta-language character of logic if one establishes set theory. Therefore, the circular character is more complex. ZFC and logic are not connected on the same language level but each theory is used as a meta-theory to define the other one. What is prior? Set theory of logic? It seems simply to be the case that there is no absolute basis for the establishment of mathematics. Relative to the reliability of set theory and first-order logic, no problem arises.[39]

Although the mathematical foundations were in a crisis at the beginning of the 20th century, we do not speak about a crisis in the foundations of mathematics today. In a certain sense, we believe that the foundations of mathematics on logic and set theory give us a reliable and sufficient basis for doing mathematics in sciences. (The research going on in set theory and logic are usually motivated by other problems, such as what consistency principles do we need for ordinary mathematics or what logics do we need in order to model certain phenomena?) Practically it is clear that this is quite reasonable simply because of the fact that it works, and we do not have an idea of a better foundation of mathematics.

What can be said about philosophical explanations? We want to argue for the thesis that under certain circumstances it is quite reasonable to accept philosophical explanations (and definitions) that are circular in nature. To make this clearer: it is indisputable that the ordinary way to give an analysis of philosophical contexts should be formulated in a well-founded way. But it is

[39] It should be mentioned that the content of axioms in set theory, expressing the very basic properties of sets is questionable. Some researchers think that the power set axiom is not a good axiom, other question the reliability of the union axiom.

not acceptable that a philosophical argument should be rejected simply because there is a circular aspect in the consideration. We will see in the following subsections that there are in fact philosophical analyses that are circular and at the same time reliable in their content. We begin with some remarks concerning the ontological status of events.

2.5.3 Events

We shall consider a hypothetical example. Two highly discussed (and controversial) concepts in philosophy and linguistics are the concept 'event' and (sometimes connected with it) the concept 'causality' (or 'causal relation' etc.) We consider the following two definitions of an event and a causal relation.

> Definition 1: An event is the cause or the effect of a causal connection.
>
> Definition 2: A causal connection between two events A and B is a relation that holds between A and B and is governed by regularities.

The problem is that the two definitions are circular in the sense that Definition 1 makes no sense without Definition 2 and Definition 2 makes no sense without Definition 1, provided we do not know a priori one of the concepts. Nevertheless, it is important to notice that the above definitions are not trivial in character, because the important impact (and all the work that has to be done) must be captured in the regularities that determine the relation between the events A and B. These regularities are captured by laws of nature that are representation in standard physical theories.[40]

We have not specified anything concerning these regularities, but it is only important to get a flavor of the principal idea, namely that sometimes it can happen that it is reasonable to accept definitions like the one above. To clarify things: the claim is neither that circular definitions should be preferred, nor that we should replace well-founded definitions and explanations by circular ones. The claim is simply that circularity alone should not be the reason for rejecting the introduction of a particular concept. An important concept that works quite well in applications and yields interesting results based on a circular argumentation is better than no concept at all.

2.5.4 Circularity and Epistemology

The above claims have a similarity to some recent developments in epistemology. Whereas many philosophers claim that the development of a well-founded, non-circular, and reliable theory of knowledge is in principle impossible - and

[40] It is clear that there are other alternatives to introduce events and causality. One is to take events as primitive or intuitively given without any further explanation. Another alternative is to reduce causal relations to formal laws given by an equation. It seems to be the case that physics tends more and more to become a discipline where everything can be described by equations (or formulas).

therefore prefer one or the other form of relativism - Ernest Sosa proposes a theory of knowledge that is circular to a certain extend but nevertheless should be reliable (according to Sosa). Sosa wants to support a kind of epistemological externalism. He disagrees with authors like Davidson, Rorty, or Williams to argue for one or the other form of internalism. According to Sosa, internalism is not a good basis for or against a certain belief because of the following counterintuitive argument: If internalism was true, every belief would be equally justifiable. He accepts the claim that a complete theory and understanding of knowledge is impossible.[41] Furthermore, he agrees that there is a circular character in the concept of knowledge. The following quotation makes this clear.

> *"The answer might come back: 'But once we had an argument A for W* [where W represents a total way of forming beliefs, K.K.] *being reliable from premises already accepted, we would embed our faith in W's reliability within a more comprehensive coherent whole that would include the premises of our argument A.' And it must be granted that such an argument would bring that benefit. However: we know that such an argument would have to be epistemically circular, since its premises can only qualify as beliefs of ours through the use of way W. That is to say, a correct and full response to rational pressure for disclosure of what justifies one in upholding the premises must circle back down to the truth of the conclusion."*[42]

We can describe this situation as follows: Even if we have a 'total way of forming beliefs W' - or a theory W that provides us with a methodology for forming beliefs - this very theory itself is subject to a theory of knowledge, because it is not immediately given (if one argues for a rationalist position). In order to justify the premises for that theory W, we have to refer back to W. That is essentially a circular move. In other words: In order to judge whether a belief is reliable or not, we have to refer to W in order to establish a basis to form beliefs. This very theory itself must be justified as knowledge (or belief), a task that cannot be done in a non-circular way. But as Sosa points out, this is not a breakdown argument for the very idea of a theory of knowledge. The following quotation makes this clear.

> *"E: A belief B in a general epistemological account of when beliefs are justified (or apt), when that applies to B itself and explains in virtue of what it, too, is justified (apt).*
> *G: A statement S of a general account of when statements is grammatical (...) that applies to S itself and explains in virtue of what it, too, is grammatical."*
> *P: A belief B in a general psychological account of how one acquires*

[41] Cf. [So94], p.93 and p.109.
[42] Cf. [So94], p.107.

> *the beliefs one holds, an account that applies to B itself and explains why it, too, is held.*
> *Why should E be more problematic than G or P? Why should there be any more problem for a general epistemology than there would be for a general grammar the grammaticality of whose statement is explained in turn by itself, or for a general psychology belief in which is explained by that very psychology."*[43]

Sosa's argumentation yields the claim that, although a certain reasoning might be circular (or includes a circular reference, or can be applied to itself), this reasoning cannot be necessarily rejected. This is in the spirit of this work: circular phenomena can be detected in various areas, and the task is not to eliminate them per se, but rather to accept the facts and model them as far as this is possible. Not the existence of circularity is the problem, but the reliability of the argument.

In the next section, we shall consider certain conceptual aspects of circularity. The idea is to give a conceptual explanation of what circularity is.

2.6 A Conceptual Discussion of Circularity

We considered a lot of examples of circular phenomena. Although from an intuitive perspective it is relatively clear what we mean when we talk about a circular phenomenon, it is much more difficult to make it conceptually clear (on a philosophically sound basis) what circularity is. In other words, although the described phenomena in this section are intuitively obvious, it is nevertheless an obstinate problem to define what it is for an entity to be circular.[44] This section is devoted to this problem.

It is clear that not every paradoxical situation is necessarily circular, and not every circular situation is necessarily paradoxical. We saw in the sections above that there are a lot of phenomena that are circular in nature but make perfectly good sense in ordinary (intuitive) reasoning. Conceptual examinations of circularity need to address this property and need to give reasons why this is the case.

Standardly, it is supposed that the essence of circularity lies in its property to be self-reflexive: For example, the Liar sentence is circular, because this sentence contains a demonstrative pronoun that refers to the sentence itself, not an extrinsic entity. Most of the classical examples of circular phenomena point this out. Additionally, most of the classical definitions of circularity are based on this idea. We state this as a first approximation of a conceptual definition of circular phenomena.

[43] Cf. [So94], pp.110/111.

[44] We use the notion *entity* for anything that can be circular (for example, sentences, propositions, predicates, arguemnts etc.).

> (1) *An entity is circular if and only if it refers in one or the other way to itself.*

Although this condition is easy to apply to different examples, it does not seem to be the correct condition for a conceptual characterization of circularity. Whereas self-reference is clearly a sufficient condition for circularity, it is not a necessary condition. The Liar circle (5) in Section 2.2 is a counter-example. No sentence in (5) refers to itself and nevertheless (5) is circular. It makes simply no sense to say that the collection of all sentences in (5) refer together again to the collection of sentences in (5). The reason for this is the fact that it is not clear what it means for a collection of sentences to refer to itself. The collection does not refer to the collection of all sentences, but every sentence refers to another sentence in the collection. Therefore, it is not possible to keep (1) as a good definition.

If we take the Liar circle into account on the one hand, and the fact that self-reference is nothing else than a circle of length one on the other hand, we could generalize the idea by saying that circularity consists in the fact that there is a circle of length n (for $n \in \mathbb{N}$) concerning the reference of the entity. Although this explanation is itself circular in character, it comes quite close to the state of affairs. Graphically, it is possible to draw a diagram for the reference. We state this idea as a second (and hopefully better) approximation of circularity.

> (2) *An entity is circular if and only if the chain of references of the entity finally refers back to the entity itself.*

Notice that the expression *"chain of reference ... refers back to itself"* essentially describes the features of a circle. In a certain sense, we do not explain circularity in a classical sense: to say that circularity is essentially to have a circle in the reference chain does not satisfy anyone who is sceptical concerning circular explanations, because we are explaining the definiendum with itself (or with a slight simplification of the definiendum). Additionally, (2) does not explain why there are good forms of circularity and why there are bad forms of circularity. Why does circularity sometimes lead to paradoxes and sometimes does not? How can we fix this problem?

The common feature of the above approximations of a conceptual definition of circularity is the fact that self-reference as well as circle reference of length n yield an infinite regress. This infinite regress can be described as a non-well-founded phenomenon. This seems to be an important point, because nearly all kinds of cognitive achievements can be summarized as follows: try to find a basis on which the whole reasoning can be founded. If this basis is established use rules of logic (clearly these rules can vary quite significantly dependent on the ideological background and the topic of the theory) or common sense reasoning in order to develop a theory for a phenomenon. The similarity to inductive reasoning is obvious. Human reasoning is most often a representation of the world where reasoning is based on assumptions taken to be intuitively correct and developed further via (intuitively) correct rules of logic. We can state the well-foundedness hypothesis as a further approximation for a definition of circularity.

> (3) *An entity is circular if and only if the entity is non-well-founded. Well-foundedness means that there is a base case to start an evaluation and there are rules that govern the further steps.*

Many examples fit quite nicely into this conceptual definition of circularity. Unfortunately, a similar problem as in the above approximation (2) arises: there are circles where the collection of sentences is not itself non-well-founded, but every single sentence is non-well-founded. To describe the feature of circular entities to be not well-founded more closely we can say that at least there is one aspect (which can be a truth value, a semantical evaluation, common sense reasoning, or anything else) of the entity that is non-well-founded. In other words, the problem of the Liar sentence is not that there is an intrinsic self-reference in it, but that there is no possibility to assign a well-founded truth value of this sentence that is inductively based on the assignment function assigning truth values to terms. Or again expressed in other words: the problem of the Liar circle is that there is no way to assign a well-founded truth value to any of the sentences, precisely because the same thing as in the case of the classical Liar sentence happens. Notice that we do not claim that the collection of all sentences of the Liar circle is non-well-founded.[45] But we do claim that there are aspects in the Liar circle that include non-well-foundedness. We give the fourth approximation of a definition of circularity as follows.

> (4) An entity is circular if and only if there are aspects of this entity that are non-well-founded.[46]

Can Definition (4) explain why some forms of circularity yield paradoxes (or can cause a pathological behavior in general) but others do not? Obviously this is not the case. Up to now, there is no possibility to have criteria in order to make such distinctions. What can be said about this problem? Well, in a certain sense, this whole work tries to give an answer to this question. In a mathematical context, a circularly defined object is pathological if it is impossible to define it using well-founded principles.[47] In Part III of this work, we will see that it is possible to define subsets of ω circularly that are also definable using classical (second-order) techniques of mathematics.

We can say the following. A circular phenomena is pathological in its behavior if there is no well-founded possibility to define it. On the other hand,

[45] It seems hard to get an idea of what that means at all.

[46] We take the term *aspect* as an intuitively given concept that does not need further explanations.

[47] This claim is stronger than the claim that every object needs to be defined by an induction. A pathological circular definition means in this context a definition that defines an object outside of mathematics.

phenomena that do not behave paradoxically do allow in one or the other way a well-founded representation (modeling).[48] We state this as a conjecture.

Conjecture 2.6.1 *A circular entity is pathological if it does not allow an appropriate representation that is well-founded in the mathematical sense of well-foundedness.*[49]

This conjecture has the features of a philosophical thesis that must be examined more closely with the means of mathematical tools. This work can be interpreted as attempt to support this thesis. Although we do not think that this work will give us a complete overview and a complete discussion of this topic, at least one can consider it as a step towards a philosophically sound examination of the nature of circularity.

2.7 History

Semantic paradoxes like the Liar sentence and the problem of truth in natural languages are well-known since ancient times, although there are not many written texts concerned with this topic in the classical Greek and Roman tradition. The example of the Cretan, who claims that all Cretans are liars, was first mentioned by Epimenides. Diogenes refers to Eubulides as one of the first who discovered Liar-like sentences. Furthermore, Aristotle, Theophrastus, and Chrysippus can count as ancient writers about the Liar paraodx.[50] (It is believed that Theophrastus has written ten books about the Liar paradox.)[51] Different versions of semantic paradoxes in the medieval times can be found in Buridan (cf. [Bu82]), a medieval philosopher who wrote a book about paradoxes and provided a certain solution for the Liar paradox by claiming that the Liar-like sentences are plainly false. A good collection of semantic paradoxes can be found in [BarEt87] and [GuBe93]. A further reference is [Mc91].

The father of the examination of truth concepts in formal languages is Alfred Tarski. He was the one who proved in [Ta56] that a truth predicate is not definable in the object language in sufficiently strong languages. Circularity in the theory of knowledge was (as far as the author knows) first discovered by David Lewis in his Ph.D. thesis [Le69]. A further analysis and the modeling of common ground versus private knowledge were provided by Barwise (Cf. [Bar90]). Further ideas concerning this topic can be found in [BarMo96] and for a different formal account in [BaMoSo∞].

A good overview of circularity in mathematics, computer sciences, philosophy and other fields is provided by [BarMo96]. Probably, this is the only available book that was written in an interdisciplinary style including many different features of circularity. Circularity in philosophical argumentation is

[48] Notice that this includes the definition using transfinite inductions.

[49] Notice that the mathematical sense of well-foundedness is clearly and precisely specified. It is not necessary to add further specifications concerning the nature of mathematical well-foundedness.

[50] For interesting historical remarks concerning ancient books about the Liar paradox we refer the reader to [Bo65], p.151 and pp.250ff.

[51] Compare [Bo65] for further information concerning the historical context of paradoxes.

probably as old as philosophy. It is quite common in philosophical discourses to claim that a particular argument is circular in order to reject it.

Chapter 3

An Overview of the Topic

The literature about circularity and in particular about special techniques to model circularity is enormous. It is not easy to see the trees in the woods because of the overwhelming number of papers and books about this topic. As a consequence a work discussing a collection of different approaches to circularity must be necessarily incomplete. This dissertation is not an exception. In this chapter, we summarize which topics are captured in this work and we give some reasons why the material is chosen in this way.

We begin with the most famous example of circular phenomena, namely the pathological behavior of circular sentences in sufficiently strong languages. If one takes semantic paradoxes, in particular, the Liar paradox seriously, then there are different possibilities to model circularity and there are different possibilities simply to block paradoxes in one or the other way. These two possibilities are the choices one can adopt in order to prevent the whole system from becoming inconsistent. Because of Tarski's theorem, it is impossible to define a truth predicate in sufficiently strong languages. First, we consider some strategies to block pathological sentences.

One can restrict the syntax of the object language. This was done by Tarski himself and yields the introduction of a hierarchy of languages.[1] Russell and Whitehead (in [RuWh27]) also used type theory to prevent their system from facing the same destiny as Frege's system, namely from becoming inconsistent. Intuitively, this strategy shifts expressions about the truth of sentences to another language.

If one considers semantical paradoxes in natural languages, the strategy to block circular sentences can be translated into the statement as the Liar sentence is syntactically non-well-formed (in the object language). Although this seems to be counterintuitive - because obviously there is nothing syntactically wrong with a Liar-like sentence - this is at least a possibility to solve the problem with brute force. Notice that even in formal languages there are no reasons at all why we should argue that a Liar-like sentence of arithmetic cannot be expressed. The advantage of this strategy is to preserve classical logic and classical model theory.

A little bit more sophisticated concerning the syntactical restrictions of lan-

[1] Cf. [Ta56].

guages is the strategy to argue that self-referential sentences are syntactically well-formed sentences, but they do not express propositions. As a consequence, Liar-like sentences are sentences, but they cannot produce paradoxes, because they do not denote propositions that can be semantically evaluated.[2] This account is quite old and was historically proposed by John Buridanus.[3]

Although the last strategy seems to be quite intuitive, there are also doubts whether this is a reasonable solution. First, in the case we reason about the Liar sentence we reason in a particular way. For example, consider the following reasoning: "Assume the Liar sentence is true, then it must be true what the Liar sentence claims, and therefore the Liar sentence is false etc.". It is quite counterintuitive to give an explanation of what we are doing when we reason about this sentence, if we assume at the same time that we do not refer to (or express) a proposition while we are reasoning. Another problem is that this account works for Liar-like sentences, but other forms and occurrences of circularity cannot be modeled in this approach. A further reason for the claim that this path is not very interesting from our perspective is the following: how can we methodologically sound argue which self-referential sentences denote propositions and which sentences do not? We saw in Chapter 2 that there are self-referential sentences that make perfectly good sense (and are not paradoxical), and we do not want to exclude these sentences from our language. How shall we decide which sentences denote propositions and which sentences do not without begging the question?[4] A final reason for rejecting this 'simple' solution is the fact that for the development of formal semantics for natural languages, it is quite important to find a possibility to represent pathological propositions in our theory. This is not possible by simply saying there are no such propositions.

Because of these reasons nothing will be found in this work about a 'modeling' of circularity by simply rejecting that there is anything like circularity in natural or formal languages on an object language level. We want to examine more sophisticated frameworks in which there exists a possibility to formulate self-referential sentences (or propositions). These frameworks cannot be as easy as the above rejection and require sometimes a quite large amount of sophisticated mathematical framework, but the advantage is that these frameworks give us the chance for a better analysis and a better understanding of the phenomena. Moreover, they help to give a more precise idea as to what circularity really is.

We shall consider the second alternative, namely the alternative which tries to model circularity. In doing this, one has several choices.[5] Three possibilities to model circularity can be summarized as follows:

[2] It is important to notice that this move does solve the problem only if one adopts a theory where propositions are the bearers of truth.

[3] Cf. [Bu82].

[4] Notice that the choice which sentences should be excluded presupposes an analysis and understanding of the paradoxes. The described attempt is in fact a form of begging the question.

[5] Compare the following remarks with [Fe84].

- Change the underlying logic.
- Change the semantics (model theory).
- Change the basic objects of mathematical discourse, i.e. introduce circular sets.

The first possibility mentioned above prevents paradoxes by modifying the underlying logic. This results directly in one or the other kind of non-classical logic. For example, Kripke's approach[6] uses a non-classical logic with three truth values (a sentence can be true, false, or neither true nor false) instead of the classical bivalent logic. Because logicians have developed an incredible number of different non-classical logics in the last sixty years, there are many possibilities to reformulate Kripke's approach in other logics as well.

What else can be said about the logical account of modeling circularity? The modification of the underlying logic gives us a possibility for a formal representation of Liar-like sentences in the object language in a consistent framework. In non-classical logic, not every sentence must necessarily be true or false. Therefore, we can have pathological sentences in our language (like the Liar sentence) and still be consistent because these sentences get a non-classical truth value (for example 'undefined', or 'neither true nor false', or 'both true and false' etc.). We will consider this branch of modeling circularity concerning the semantical paradoxes in Part II of this work. We shall use the quite general framework of bilattices in order to represent the ideas of Kripke and Martin/Woodruff.[7] Using bilattices (instead of semilattices as Kripke proposed) has the advantage to be more general and to be able to model different approaches in one framework. Bilattices will give us a tool at hand that is general enough to model the three-valued Kripke account (using the strong-Kleene evaluation), the three-valued Priest account using a paraconsistent logic,[8] and the four-valued Visser account[9] as special cases of our general development. The generality of the bilattice approach justifies the algebraically more difficult structure of a bilattice in comparison with the structures used in ordinary presentation of Kripke's theory, usually formulated using CCPOs (so-called coherent complete partial orders) or semilattices.

Partially defined truth predicates in a non-classical logic have some negative features. One problem with this account is that it is not clear how to extend the idea of a partial logic to other cases of circularity that are different from the problem of truth. We saw in Chapter 2 that there are occurrences of circularity in various disciplines. It is not clear how the partial logic approach can be straightforwardly extended to these other phenomena. This is a feature that weakens this particular idea quite strongly, if one takes a general perspective into account. Nevertheless, the idea of partial logics to model semantical paradoxes will be considered in Part II of this work as an important development towards an appropriate theory of circularity. We will discuss the

[6] Cf. [Kr75].

[7] Cf. [Kr75] and [MaWo75].

[8] Cf. [Pr79].

[9] Cf. [Vi84].

problems of how to extend this approach to other applications as well in Part II.

A further possibility we mentioned to model circular phenomena is to adopt an alternative semantics. The idea is that syntactically self-referential sentences are allowed in the object language and at the same time we use a classical two-valued logic in order to evalute the sentences. The difference is to change the model theory in a way that enlarges the expressive power of the theory but that does not make the theory itself inconsistent. This is the account Gupta and Belnap use in order to develop the so-called revision theories (of truth).[10]

A strategy of how to reduce the complexity of second-order logic to first-order logic was developed by Henkin.[11] His idea was to introduce second-order objects as objects that can be divided in various classes. These classes can be interpreted in a first-order many-sorted logic. In this respect, Henkin did not change the logic (in the sense that he used a non-classical logic as in the Kripke account), but he changed the interpretation of the predicates in question, i.e. Henkin modified the semantics. In a certain sense, revision theories are based on a very similar idea. We will see in Part III how to extend ordinary definition theories with circular definitions without becoming inconsistent using a modification of the underlying semantics (model theory). The idea of the revision theoretic account is to treat ordinary predicates differently in comparison with predicates that are defined by a circular definition. This is a very similar idea to the account of Henkin: we introduce two different sorts of predicates.[12] Because there are many possibilities to extend classical semantics to a semantics which can consistently interpret circular definitions, we have to deal with infinitely many different semantical systems.

In Part III, we will examine revision theories and their various properties. Most prominently, we will consider the definitional strength of these systems. Because revision theories are a relatively flexible theory to model circular phenomena, a number of applications are mentioned in Chapter 9. Finally, we will discuss certain problems of revision theories and their conceptual and ontological presuppositions. Although revision theories are a strong tool to model circularity many conceivable applications of that theory are not spelled out yet.

There is a further possibility that does not fit directly into the mentioned strategies. If one models circularity, a natural idea is to change the 'nature' of the corresponding mathematical objects. Because in standard mathematics as well as in most applications of mathematical theories we are using sets in order to represent objects of the domain, the new strategy is to change the underlying set theory. This can be done using the theory of hypersets, a set theory that does not require that sets are necessarily well-founded in order to

[10] Cf. [GuBe93]. We will use quite often the notion *Gupta-Belnap systems* in order to refer to revision theories. We think that the notion Gupta-Belnap system makes the connection to semantical systems more explicit in comparison to the notion revision theory.

[11] Cf. [He50].

[12] As it turns out in Chapter 8 the Gupta-Belnap $\mathbf{S}^{\#}$ and $\mathbf{S}^{*}$ are too strong to find a recursive axiomatization.

count as respectable entities of mathematical discourse.

In classical set theory, there does not exist a set that contains itself because such a set contradicts the so-called foundation axiom of ZFC.[13] Nevertheless there is the possibility to extend the set theoretical universe with circular sets (sets that contain themselves), such that the resulting set theory is consistent relative to the consistency of ZFC. This extension of the universe of sets can be done by dropping the foundation axiom and adding instead of the foundation axiom one or the other form of a so-called anti-foundation axiom. We will see in Part IV how this can be accomplished. In non-well-founded set theory (sometimes also called the theory of hypersets), we can state an analogous theorem to the recursion theorem in ZFC. This corresponding theorem is called corecursion theorem.[14] This will lead us to some further considerations in category theory and the theory of coalgebras. Recently, these theories became more and more important for theoretical computer science. Another path of the theory of hypersets can be seen in applications: mostly, we will work in one or the other form of situation theory. In situation theory, it is possible to represent Liar-like propositions as well as the differences between common ground and private knowledge. We will consider these aspects in Chapter 15.

Working in situation theory will support the fact that circularity is highly context (or situation) dependent. In the previous accounts for circularity (Kripke's account as well as revision theories), contexts play no role in the modeling of circularity. In situation theory, contexts and constraints can be modeled to a certain extend, but situation theory is not general enough to give a sufficient framework for all phenomena. The reason for this is, because situation theory has no possibility to speak about constraints in the model. We will add some more comments concerning this point in the final remarks (Chapter 16) where we will consider some basic definitions from channel theory in order to get an idea of what a possible further extension could be.

Our overview is very general in nature, because it categorizes everything in relatively coarse concepts. There is a great number of works dealing with circularity and paradoxes and it does not seem to be possible all the time to find a particular account in our categorization. Our choice to consider the most developed theories in each category is a natural choice that is justified by the empirical results of these theories.

[13] For further information concerning ZFC set theory and their properties the reader is referred to Chapter 11.

[14] Originally, Peter Aczel called this theorem the final coalgebra theorem, because it is based on the properties of the final coalgebra (of a given category) as well as the recursion theorem is based on the properties of the the initial algebra (of a given category). Compare Chapter 14 for more information.

Part II

Partiality of the Underlying Logic: Kripke's Account in the Generalized Setting of Interlaced Bilattices

Chapter 4

Kripke's Theory of Truth

Since the very influential work of Alfred Tarski on the non-definability of a truth predicate in sufficiently strong languages,[1] many researchers tried to develop a concept of truth that is able to circumvent the non-definability problem. A major breakthrough was reached by the idea that truth predicates can be defined partially (using a non-classical logic). These ideas were independently formulated by Saul Kripke (Cf. [Kr75]) and Robert Martin and Peter Woodruff (Cf. [MaWo75]) at approximately the same time.[2]

The idea of Kripke's account is to use Tarski's biconditionals in order to get a transfinite hierarchy of new extensions of the truth predicate. Because classical logic is not monotone, there is no chance to reach a fixed point in this process using classical logic.[3] The idea of Kripke (and Martin and Woodruff as well) was (were) to use a monotone non-classical logic. Whereas Kripke used the so-called three-valued strong Kleene logic, Martin and Woodruff worked in the weak counterpart, namely the three-valued weak Kleene logic. The main difference between these logics is the logical evaluation of formulas including subformulas that are assigned the third truth value. A good reference for these logics is [Kl52].

In this chapter, we will introduce Kripke's theory of partially defined truth predicates on an informal level. In Section 4.1, we will present the basic idea of Kripke's account as well as the background theory Tarski developed. Section 4.2 deals with the algebraic counterparts of certain non-classical logics (three-valued and four-valued logics): so-called CCPOs function as the algebraic counterpart of three-valued logic and interlaced bilattices as the counterpart of four-valued logic. We will summarize some of the most important properties of certain algebraic structures. Section 4.3 deals with a generalized version of Kripke's original construction that finally yields the desired result, namely that truth

[1] Cf. [Ta56].

[2] Although Kripke and Martin/Woodruff deserve equal credit for developing the ideas that will be explained in this chapter, we will refer to this account by the term 'Kripke's account'. This is common in the literature. We do not intend to diminish or underestimate the credit of Martin and Woodruff in any respect.

[3] It should be mentioned that it is possible to define truth using classical logic. The idea is to restrict the theory in other respects, for example avoiding unrestricted quantification, or using only functions of a limited sort. More information concerning these theories can be found in [Sc75] and [Fe84].

predicates can be defined in the object language. In Section 4.4, we shall discuss some of the consequences of partially defined truth predicates.

4.1 Partially Defined Truth Predicates

4.1.1 Tarski's account

It was Tarski who recognized that a truth predicate for languages should satisfy a certain condition, in order to have an intuitively correct behavior. This condition is commonly called Tarski's 'condition T'. We can formulate this concept as follows (provided that a language L is given and ϕ is a sentence of L):

$$(1) \quad ”\phi” \quad \text{is true} \;\Leftrightarrow\; \phi$$

In (1), ”ϕ” is a name (or code) for the sentence ϕ of the object language. In other words, (1) expresses the relation that the truth predicate applied to the name (code) of ϕ holds if and only if ϕ holds as well (relative to a given model). Notice that (1) itself is not a statement in the object language, because the equivalence relation '$\Leftrightarrow$' is a symbol of the meta-language (that means the language we are using to reason about the object language). The left side as well as the right side of (1) are formulas of the object language.

Tarski's new insight was that even though his condition T is intuitively plausible for a truth theory of arbitrary languages, contradictions are deducible if one tries to define such a predicate in a sufficiently strong language. To show this result assume $\mathbf{T}(\phi)$ is an abbreviation for the expression ”${}”\phi”$ is true”. Consider a language L in which a sentence ϕ with the property $\neg\mathbf{T}(\phi) \Leftrightarrow \phi$ exists.[4] If L can code arithmetic (for practical purposes one can have in mind ordinary Peano Arithmetic), then such a sentence exists, because of the Diagonal Lemma.[5] To prove that a contradiction can be deduced we reason as follows:

$$\begin{array}{lll}
(1) & \neg\mathbf{T}(\phi) \Leftrightarrow \phi & \text{assumption} \\
(2) & \neg\mathbf{T}(\phi) \Leftrightarrow \phi \text{ and } \phi \Leftrightarrow \mathbf{T}(\phi) & \text{(1) and definition of } \mathbf{T} \\
(3) & \neg\mathbf{T}(\phi) \Leftrightarrow \mathbf{T}(\phi) & \text{(2) and transitivity of ”}\Leftrightarrow\text{”} \\
(4) & (\neg\mathbf{T}(\phi) \Rightarrow \mathbf{T}(\phi)) \wedge (\mathbf{T}(\phi) \Rightarrow \neg\mathbf{T}(\phi)) & \text{(3) and logic} \\
(5) & \mathbf{T}(\phi) \Rightarrow \neg\mathbf{T}(\phi) & \text{(4) and logic} \\
(6) & (\mathbf{T}(\phi) \Rightarrow \neg\mathbf{T}(\phi)) \Rightarrow \neg\mathbf{T}(\phi) & \text{logic} \\
(7) & \neg\mathbf{T}(\phi) & \text{Modus ponens of (5) and (6)} \\
(8) & \neg\mathbf{T}(\phi) \Rightarrow \mathbf{T}(\phi) & \text{(4)} \\
(9) & \mathbf{T}(\phi) & \text{Modus ponens of (7) and (8)}
\end{array}$$

Lines (7) and (9) are inconsistent and we deduced a contradiction. Notice that the reasoning we are using to deduce this contradiction is based on classi-

[4] The existence of this sentence can only be expressed, if the considered language is strong enough. That means that L contains an elementary coding scheme, for example, standard arithmetic.

[5] Cf. [Ob93] or [Sm92] for readable presentations of this lemma.

cal logic plus 'condition T' as the definitional property of the predicate **T**. The existence of this proof is not dependent on special rules of the deduction calculus, because we simply used Modus Ponens, a deduction rule that is included in every logical calculus.

As a solution for the above problem, Tarski introduced a hierarchy of languages in order to define truth predicates for the languages that are lower in that hierarchy. Although this seems to be a reasonable way to get a solution for the problem, this does not work for natural language, because in natural language we have no possibility to define a truth predicate in a meta-language. We have only, say, English, but not an additional language, say 'Meta-English'. Additionally, natural language includes paradoxes, like the Liar sentence. Tarski's account does not model these paradoxes. His account plainly declares that such sentences are syntactically ill-formed (in the object language). His strategy suffices to block pathological behavior, but his account cannot model the different types of pathological phenomena. Therefore, the described strategy does not lead us to a point where we can use it reasonably for our purposes. Kripke's account tries to create a framework that prevents paradoxes but nevertheless allows Liar-like sentences. The next subsection gives an idea of Kripke's approach.

4.1.2 The Idea of Kripke's Approach

Because of the described problems of Tarski's theory, Kripke introduced a new idea. Roughly speaking, he used a partial logic in order to define partial truth predicates in the object language. In order to get an intuitive understanding of Kripke's account, we will sketch the basic idea in this subsection. Notice that even Kripke's account includes essentially stages. We are not so far away from Tarski, as it seems to be.

Assume a classical ground model $\mathfrak{M}$ is given that provides interpretations of all sentences of a given language L where L does not contain a truth predicate. Now we add syntactically a truth predicate **T** to L. We define the expanded language $\mathbf{L}^+ = L \cup \{\mathbf{T}\}$ syntactically as the smallest set, such that it holds: if ϕ is a sentence of $\mathbf{L}^+$, then $\mathbf{T}(\phi)$ is also a sentence of $\mathbf{L}^+$. How can we introduce an interpretation (extension) of the new predicate **T** using stages? We can apply a transfinite induction as can be seen in the following explanations.

> Zero stage: The sentences of $\mathbf{L}^+$ that do not include a truth predicate are interpreted by the classical ground model $\mathfrak{M}$. Notice that even on this stage we get a partial extension of **T**. All sentences that are true in $\mathfrak{M}$ are in the extension of **T**. All other sentences are in the anti-extension of **T**. These are the sentences that are either false in $\mathfrak{M}$ or sentences that contain the truth predicate **T** and therefore, cannot be evaluated on stage 0.
>
> First stage: Consider a sentence of the form $\mathbf{T}(\phi)$ where $\phi \in L$.[6]

[6] To simplify the presentation we suppress the coding machinery here. Clearly, the truth predicate is only defined on codes of formulas and not on formulas themselves.

We know that ϕ can be interpreted using the ground model $\mathfrak{M}$. Now we can apply Tarski's biconditionals: $\mathbf{T}(\phi) \leftrightarrow \phi$. If ϕ is true in the ground model $\mathfrak{M}$, then $\mathbf{T}(\phi)$ is in the extension of $\mathbf{T}$. All other sentences are in the anti-extension of $\mathbf{T}$. Notice that in particular the liar sentence $\phi = \neg\mathbf{T}(\phi)$ is in the anti-extension of $\mathbf{T}$. Logically, the Liar sentence gets the truth value 'undefined'. This is crucial, because in the case that the Liar sentence would be defined as false the next stage would change the truth value to true.

Second stage: On the second stage we repeat the procedure above. All sentences of the form $\mathbf{T}(\phi)$ where ϕ is interpreted on the first stage can be interpreted using Tarski's biconditionals. For example, a sentence of the form $\mathbf{T}(\mathbf{T}(\phi))$ can be evaluated. Still it is not possible to evaluate a sentence like the Liar sentence. The extension of $\mathbf{T}$ on the second stage is the extension of $\mathbf{T}$ on the first stage plus the newly interpreted sentences.

Third stage: We apply the same procedure as above.

$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$

ω-th stage: Let us define $\mathbf{T}^n(\phi)$ as the n-th iteration of the truth predicate (for $n \in \omega$). In other words, ϕ is embedded in a sentence in the scopus of n iterated truth predicates $\mathbf{T}$. How can we define the extension of $\mathbf{T}$ on stage ω? The idea is to take the union over all true sentences in which the truth predicate is finitely many times iterated. Technically that means:

> $\mathbf{T}^\omega(\phi)$ is in the extension of $\mathbf{T}$ if and only if $\forall n < \omega : \mathbf{T}^n(\phi)$ is in the extension of $\mathbf{T}$.

$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$

λ-th stage where λ is a limit ordinal: We define the extension of $\mathbf{T}$ in a similar way as defined on the ω-th stage. Formally we get:

> $\mathbf{T}^\lambda(\phi)$ is in the extention of $\mathbf{T}$ if and only if $\forall \alpha < \lambda : \mathbf{T}^\alpha(\phi)$ is in the extension of $\mathbf{T}$.

Remark 4.1.1 (i) Notice that at no stage the Liar sentence will be interpreted. It cannot be evaluated on a successor level, because there is no predecessor stage where this sentence can be evaluated. Furthermore, it can never be true even on a limit stage λ, because the Liar sentence is never evaluated on a stage smaller than λ. In other words, the Liar sentence will remain in the antiextension of $\mathbf{T}$.

(ii) The stage-by-stage definition of the extension of $\mathbf{T}$ is monotone in the following sense. If a sentence ϕ of $\mathbf{L}^+$ is evaluated on a stage α, then ϕ will be evaluated as true on stage $\alpha + \beta$ for every $\beta \in ORD$.[7] The same is true for every false sentence. In other words: all true sentences remain true on every higher stage, whereas all false sentences remain false on every higher stage. This implies that the extension of the truth predicate $\mathbf{T}$ increases on every further stage (in the sense that it increases or remains equal.) Precisely here the monotonicity condition of the three-valued logic comes in. Notice that in classical logic it is impossible to assign a truth value *neither true nor false* to a sentence.

(iii) With the above remark, one can show using set theoretical arguments concerning monotone operators on sets that $\mathbf{T}$ must have a fixed point: every partial extension of the truth predicate $\mathbf{T}$ is a set. Every monotone operator on sets has a fixed point. Therefore, there is an ordinal β, such that the extension of $\mathbf{T}$ on the β-th stage is equal to the extension of $\mathbf{T}$ on the $\beta + 1$-st stage. Then, the extension remains fixed on every $\beta + \beta'$-th stage, where β' is an arbitrary ordinal.

(iv) What is wrong, if one works in a classical (two-valued) logic and not in a partial logic (for example a three-valued logic)? On the zero-stage, the Liar sentence cannot be evaluated as true, therefore this sentence gets the truth value false.[8] Because of the pathological behavior of the Liar sentence it must be evaluated as true on the following stage. Then, the Liar sentence is in the extension of the truth predicate $\mathbf{T}$ on the first stage. On the second stage, the Liar sentence changes to the anti-extension and so on. As a consequence, there cannot be a fixed point determining the extension of the truth predicate. Conclude: the usage of a monotone (partial) logic is crucial in Kripke's approach.

(v) A remark concerning the terminology of Kripke is helpful. Kripke calls a sentence ϕ grounded, if ϕ is interpreted on some stage. Notice that on every higher stage, the truth value of ϕ remains fixed, i.e. ϕ does not change its truth value. Because of this property the evaluation is monotone. One can rephrase this fact as follows: in the minimal fixed point, ϕ has a definite truth value. This does not exclude the possibility that some sentences have no definite truth values in the minimal fixed point, but they have truth values in a maximal fixed point. For example, it turns out that the Truth-teller has no definite truth value (true or false) in the minimal fixed point, but it does have a definite truth value in some maximal fixed point. It is possible that in one maximal fixed point the Truth-teller is true in another maximal fixed point the Truth-teller is false. We will add some remarks of these dependencies, and the existence of minimal and maximal fixed points later.

[7] ORD denotes the class of all ordinals.

[8] Notice that in Kripke's account this sentence gets the truth value neither true nor false.

(vi) Other important concepts besides the concept 'grounded' are the following ones: A sentence ϕ is called paradoxical, if there is no fixed point where ϕ is true or false (as an example consider the Liar sentence). A sentence ϕ is called biconsistent, if there is a fixed point where ϕ is true and there is another fixed point where ϕ is false (the standard example is the Truth-teller).

In many respects, we were not very precise in developing Kripke's approach. Kripke developed his ideas in the framework of semilattices. Later it became clear that the ultimate structure in question for a three-valued Kripke account are CCPOs (coherent complete partial orders), originally introduced by Barendregt[9]. We will not develop Kripke's fixed point approach in the framework of CCPOs or semilattices in all details, because there is a generalization of his approach in the framework of so-called interlaced bilattices.

4.2 CCPOs and Bilattices

4.2.1 CCPOs

First, we will introduce CCPOs (coherent complete partial orders). The order theoretic structure of Kleene's three-valued logic corresponds to a CCPO.

Definition 4.2.1 *(i) Assume $\langle D, \leq \rangle$ is a partially ordered set. A subset $X \subseteq D$ is called consistent if it holds:*

$$(\forall x \in X)(\forall y \in X)(\exists z \in D) : x \leq z \wedge y \leq z$$

(ii) Again let $\langle D, \leq \rangle$ be a partially ordered set. We call $\langle D, \leq \rangle$ a CCPO (coherent complete partial order) if $\langle D, \leq \rangle$ has a bottom element $\perp$ and every consistent subset $X \subseteq D$ has a supremum.
(iii) A partially order set $\langle D, \leq \rangle$ is called a complete lattice if every subset $X \subseteq D$ has a supremum and an infimum.

It is clear that every complete lattice is also a CCPO. The concept of a CPO (complete partial order) is better known as the concept of a CCPO, because of a long tradition of applications of CPOs in the theory of partial orders, domain theory, and computer science in general. We would like to stress the differences between CPOs and CCPOs. In order to do this, we need to define CPOs.

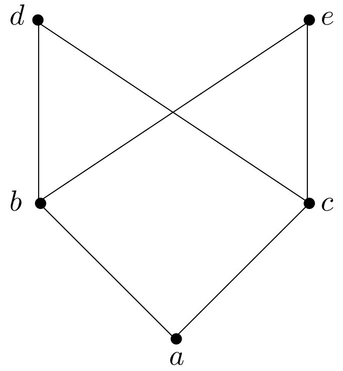
Definition 4.2.2 *A partially ordered set $\langle D, \leq \rangle$ is called a CPO, if $\langle D, \leq \rangle$ has a bottom element $\perp$ and every directed subset $X \subseteq D$ has a supremum, where $X \subseteq D$ is called directed, if it holds:*

$$(\forall x \in X)(\forall y \in X)(\exists z \in X) : x \leq z \wedge y \leq z$$

Remark 4.2.1 Although the difference in the definitions between a CCPO and a CPO seems to be quite tiny, the structures differ significantly concerning

[9] Cf. [Ba81].

their properties. For example, it is easy to check that the following structure is a CPO but not a CCPO. To see this consider the set $\{b, c\}$ and apply the definitions of CCPO and CPO. Whereas the set $\{b, c\}$ is consistent, it is not directed. Because $\sup\{b, c\}$ does not exists, the following structure is a CPO but not a CCPO.



(ii) Notice that the infinite chain $\langle \mathbb{N}, \leq \rangle$ of the natural numbers is neither a CCPO nor a CPO. If we consider $\langle \mathbb{N} \cup \{\mathbb{N}\}, \leq \rangle$ the chain of natural numbers extended by a non-standard element $\mathbb{N}$, then this structure is a CPO as well as a CCPO.

(iii) It is well-known that the class of all CPOs is equal to the class of all partially ordered sets with bottom element and the property that every chain has a supremum (chain completeness).[10] Therefore, chain completeness is the constitutive property of CPOs. But this is not sufficient for CCPOs. In the latter structures, additional properties must be satisfied.

(iv) Using elementary reasoning one can show that every monotone operator defined on a CCPO $\langle D, \leq \rangle$ has a minimal fixed point, and at least one maximal fixed point. Additionally, it is provable that the generated set of fixed points is itself a CCPO. Using these theorems Kripke proved an important fact: there exists the so-called maximal intrinsic fixed point. A maximal intrinsic fixed point x is a fixed point that is consistent with every other fixed point of the CCPO, in the sense that for every other fixed point d, $\{x, d\}$ has a supremum. The existence of such a fixed point can also be proven in CPOs (compare [Ku96a]).

Because the above construction is simply a special case in a more general setting, we will develop Kripke's account in the framework of interlaced bilattices. This kind of structure was originally introduced in theoretical computer science and is a generalization of the well-known concept of the theory of lat-

[10] Cf. [DaPr90] for further information concerning this non-trivial equivalence.

tices. We begin our closer look with the definition of a billatice according to [Gi88].

4.2.2 Bilattices

Bilattices were introduced in theoretical computer science and AI. Programs based on classical logic cannot deal without problems with underspecified situations (not enough information available) and overspecified situations (too much and inconsistent information available). A natural idea is to introduce additional truth values to the framework in order to assign a truth value to an underspecified (or overspecified) situation different from true and false. That was the starting point for Ginsberg to introduce bilattices in so-called default logic. We begin with his original definition of the concept of a bilattice as specified in [Gi88].

Definition 4.2.3 *A bilattice is a six-tuple $\langle D, \wedge, \vee, \cdot, +, \neg\rangle$, such that the following two conditions (i) and (ii) hold:*

(i) $\langle D, \wedge, \vee\rangle$ and $\langle D, \cdot, +\rangle$ are complete lattices.
(ii) $\neg : D \longrightarrow D$ is a lattice homomorphism with the properties
$\neg : \langle D, \wedge, \vee\rangle \longrightarrow \langle D, \wedge, \vee\rangle$ *and*
$\neg : \langle D, \cdot, +\rangle \longrightarrow \langle D, +, \cdot\rangle$

Remark 4.2.2 (i) In our context, the lattice homomorphism $\neg$ inverting one order relation and preserving the other one can be understood as a negation. Intuitively, $\neg$ is intended to invert the truth order and preserve the information order of the underlying algebraic structure. If we have a certain information coded as a statement (sentence or formula), the negation of that statement should not influence the degree of informational coded in it. On the other hand, if we have a true sentence and we apply $\neg$ to the sentence, then the truth value should be inverted. These properties are immediate consequences of the intuitive meaning of a negation. Notice that in the definition of a bilattice the lattice homomorphism is much more general in comparison with our intended interpretation.

(ii) An alternative definition of a bilattice $\langle D, \wedge, \vee, \cdot, +, \neg\rangle$ can be achieved as follows: a structure $\langle D, \wedge, \vee, \cdot, +, \neg\rangle = \langle D, \leq_1, \leq_2, \neg\rangle$ is a bilattice if $\langle D, \leq_1\rangle$ and $\langle D, \leq_2\rangle$ are complete lattices and (a) and (b) hold:

(a) $\forall x, y \in D : x \leq_1 y \;\leftrightarrow\; \neg x \leq_1 \neg y$
(b) $\forall x, y \in D : x \leq_2 y \;\leftrightarrow\; \neg y \leq_2 \neg x$

It is easy to see that both definitions are equivalent. The difference is simply to interpret a lattice as an order theoretic structure on the one hand, and as an algebraic structure on the other hand. It is well-known that both interpretations are equivalent.

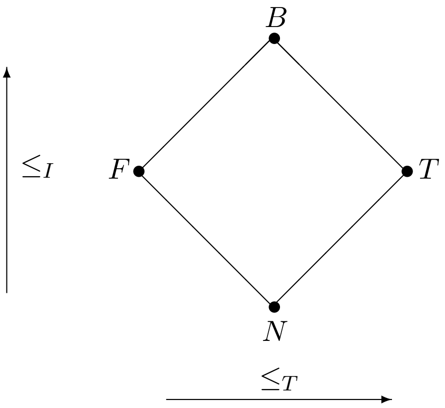
(iii) Intuitively, a bilattice is a lattice that can be read from left to the right and from the bottom to the top (provided the bilattice is finite). Notice that the definition of a bilattice is completely general and abstract. The diagrammatic representation is only a simplification to get an intuitive idea of an abstract concept. As far as the author knows it is an open problem whether any finite bilattice can be represented (in principal) diagrammatically as a 'double-Hasse-diagram' or not. For the topics presented in this chapter an answer of this question is not important.

(iv) Definition 4.2.3 is the original definition proposed by Ginsberg in [Gi86] and in [Gi88]. We need a slightly different definition of bilattice for our construction. We will define this type of bilattice below.

The following three examples should give a flavor which kind of structures bilattices are.

Example 4.2.3 (i) The trivial example of a bilattice is the structure $\langle \{x\}, \wedge, \vee, \cdot, +, \neg \rangle$ where the domain consists only of a single element x. The lattice homomorphism $\neg$ is the identity.

(ii) Consider the structure FOUR $= \langle \{N, F, T, B\}, \leq_I, \leq_T, \neg \rangle$ with four truth values where $\leq_T$ is interpreted as truth order and $\leq_I$ as information order. The lattice homomorphism $\neg$ is a negation defined on the four truth values. The order relations are given according to the following diagramm.



Obviously, $\langle \{N, F, T, B\}, \leq_I, \leq_T, \neg \rangle$ is a bilattice. Intuitively, $\neg$ should be a special kind of negation, preserving the information order and inverting the truth order. $\neg$ should have the following properties: A statement that is neither true nor false should remain neither true nor false, if we apply $\neg$, because a negation does not change the information state concerning this very statement. If we do not know whether a sentence is true or false, we do not know whether the negation of this sentence is true or false, either. The same properties should be satisfied if we consider a statement that is overspecified, namely true and false. The only possibility how this can be achieved in FOUR is to define the lattice homomorphism as follows:

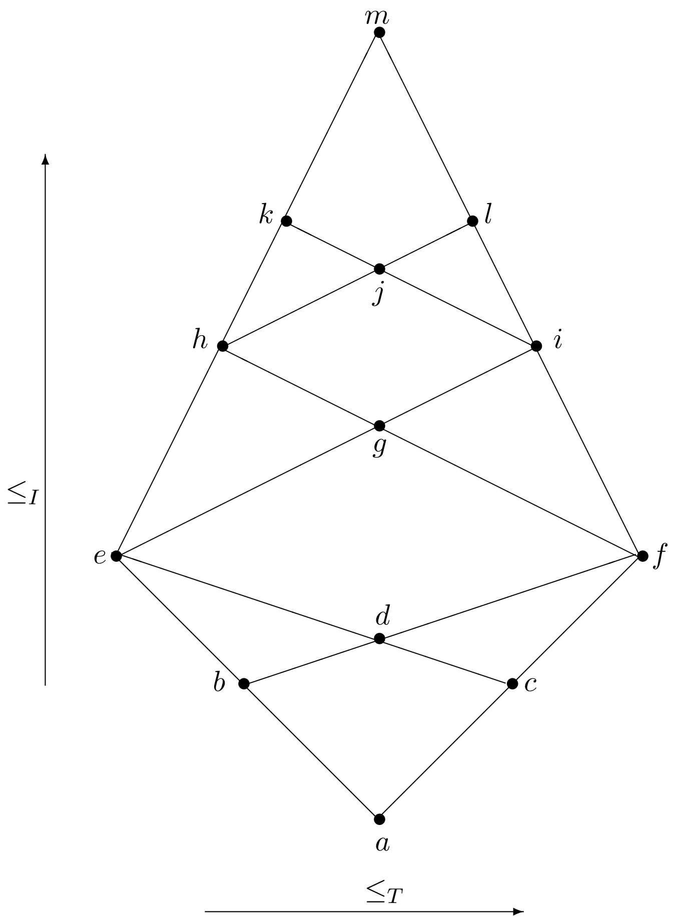
$$\neg T = F \qquad \neg F = T \qquad \neg N = N \qquad \neg B = B$$

It is quite common to identify the four truth values set theoretically with subsets of the two classical truth values {t,f}:

$$N = \emptyset \qquad T = \{t\} \qquad F = \{f\} \qquad B = \{t, f\}$$

FOUR is a generalization of the classical four-valued logics studied in other contexts, like in relevance logic or in linguistics.[11] We will not go into details concerning this point and refer the interested reader to the relevant literature.

(iii) Now, we will consider a more complicated example. The structure has the following signature: $\langle D, \leq_I, \leq_T, \neg\rangle$ where $D = \{a, b, c, d, e, f, g, h, i, j, k, l, m\}$ and the order relations are given according to the following diagram:



The lattice homomorphism ¬ preserves the nodes a, d, g, j, m of the bilattice.

[11] Cf. [Du85], [Mu89], or [Vi84].

For the other points of the domain it holds: $\neg b = c$, $\neg c = b$, $\neg e = f$, $\neg f = e$, $\neg h = i$, $\neg i = h$, $\neg k = l$, $\neg l = k$. It is easy to check that the above structure together with this homomorphism gives us a bilattice.

An easy method to construct bilattices is known from ordinary lattice theory. We use given chains to construct a bilattice. The following definition specifies the induced order relations.

Definition 4.2.4 *Assume $\mathbf{D_1} = \langle D_1, \leq_1' \rangle$ and $\mathbf{D_2} = \langle D_2, \leq_2' \rangle$ are two given chains. The product bilattice $\mathbf{D} = \langle D, \leq_1, \leq_2 \rangle$ of $\mathbf{D_1}$ and $\mathbf{D_2}$ with domain $D = D_1 \times D_2$ is specified as follows:*

$$\langle x_1, x_2 \rangle \leq_1 \langle y_1, y_2 \rangle \;\Leftrightarrow\; x_1 \leq_1' y_1 \wedge y_2 \leq_2' x_2$$

$$\langle x_1, x_2 \rangle \leq_2 \langle y_1, y_2 \rangle \;\Leftrightarrow\; x_1 \leq_1' y_1 \wedge x_2 \leq_2' y_2$$

Definition 4.2.4 provides an easy method to construct new bilattices. As we will see later, the resulting bilattices are quite restricted structures with respect to their properties.

In the following subsection, we will examine bilattices without taking into account the lattice homomorphism $\neg$. The reason for this is the fact that we will focus on order theoretical properties of bilattices and not on possibilities to introduce lattice homomorphisms. Therefore, we will denote by 'bilattice' a structure $\langle D, \leq_1, \leq_2 \rangle$ where $\langle D, \leq_1 \rangle$ and $\langle D, \leq_2 \rangle$ are complete lattices. Although it is not necessarily the case that it is possible for all these structures to introduce lattice homomorphisms with the properties of $\neg$, for many of the following examples this will be true. For more information concerning the possibilities of introducing a negation $\neg$, the reader is referred to Section 5.4.

4.2.3 Interlaced Bilattices

We need a type of bilattice that satisfies certain monotonicity conditions. This type of bilattice is called an interlaced bilattice, originally introduced by Melvin Fitting (in [Fi88]). The idea of the interlacing condition is that the supremum and the infimum operation with respect to $\leq_1$ do not only preserve the order relation $\leq_1$, but the order relation $\leq_2$ as well.[12] In other words, the calculation of the suprema and infima are monotone in both order relations. Together with distributive bilattices interlaced bilattices will be the most prominent structures in this chapter. In order to be able to define these bilattices in a very general way, we need the following definition.

Definition 4.2.5 *(i) Assume $A \subseteq D$ and $B \subseteq D$ are nonempty subsets of a given bilattice $\langle D, \leq_1, \leq_2 \rangle$. For $i \in \{1, 2\}$ the expression $A \leq_i B$ denotes the following relation:*[13]

[12] The inverse condition is also required.

[13] Cf. [Fi87, Fi89]

$$(\forall a \in A)(\exists b \in B) : a \leq_i b \ \wedge \ (\forall b \in B)(\exists a \in A) : a \leq_i b$$

(ii) Assume that $A \subseteq D$ and $B \subseteq D$ are nonempty subsets of a given bilattice $\langle D, \leq_1, \leq_2 \rangle$. For $i \in \{1,2\}$ the expression $A <_i B$ denotes the relation $A \leq_i B$ where $A \neq B$.

(iii) Assume the same conditions as above. $\bigwedge A$ denotes the infimum of A with respect to $\leq_1$ and $\bigvee A$ denotes the supremum of A with respect to $\leq_1$. ΠA denotes the infimum of A with respect to $\leq_2$, and ΣA denotes the supremum of A with respect to $\leq_2$.

It should be mentioned that for applications we have the information order $\leq_I$ and the truth order $\leq_T$ in mind. We will associate $\leq_I$ with $\leq_1$ and $\leq_T$ with $\leq_2$. Now we have the prerequisites to define the concepts of interlaced bilattices and of distributive bilattices.

Definition 4.2.6 *(i) An interlaced bilattice $\langle D, \leq_1, \leq_2 \rangle = \langle D, \wedge, \vee, \cdot, + \rangle$ is a set D which is partially ordered with respect to two partial order relations $\leq_1$ and $\leq_2$, such that $\langle D, \leq_1 \rangle$ and $\langle D, \leq_2 \rangle$ are complete lattices and the following monotonicity conditions (a)-(d) hold:*

(a) $\forall A, B \subseteq D : A \leq_1 B \ \rightarrow \ \Pi A \leq_1 \Pi B$
(b) $\forall A, B \subseteq D : A \leq_1 B \ \rightarrow \ \Sigma A \leq_1 \Sigma B$
(c) $\forall A, B \subseteq D : A \leq_2 B \ \rightarrow \ \bigwedge A \leq_2 \bigwedge B$
(d) $\forall A, B \subseteq D : A \leq_2 B \ \rightarrow \ \bigvee A \leq_2 \bigvee B$

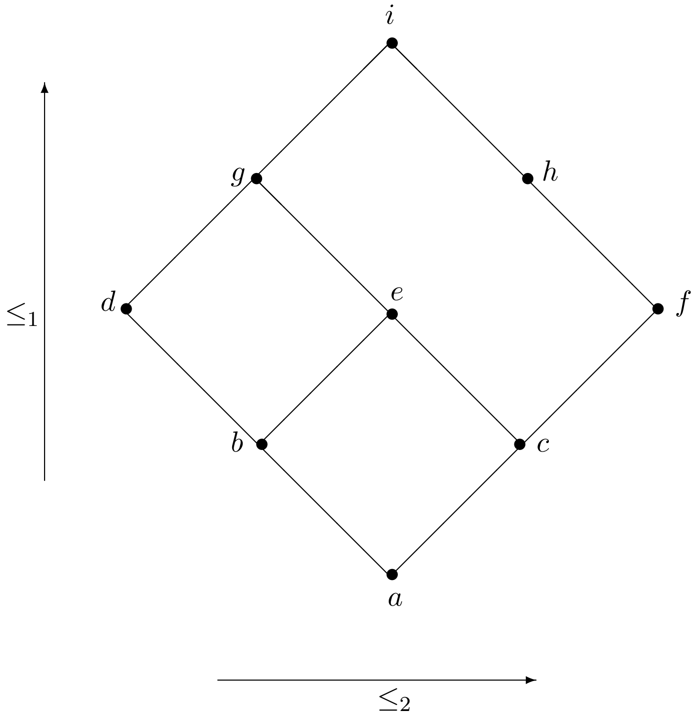
(ii) A bilattice $\langle D, \leq_1, \leq_2 \rangle = \langle D, \wedge, \vee, \cdot, + \rangle$ is called distributive, if for all $x, y, z \in D$ and $\bullet_1, \bullet_2 \in \{\wedge, \vee, \cdot, +\}$ the following relation holds:

$$x \bullet_1 (y \bullet_2 z) = (x \bullet_1 y) \bullet_2 (x \bullet_1 z)$$

In order to make the reader familiar with these new concepts, we give some examples.

Example 4.2.4 (i) Consider `FOUR`. This bilattice has the interlacing property. Additionally, this bilattice is distributive. It is the simplest non-trivial bilattice that has the interlacing property and is furthermore distributive.

(ii) Consider the following bilattice $\mathbf{D} = \langle \{a, b, c, d, e, f, g, h, i\}, \leq_1, \leq_2 \rangle$ where the order relations $\leq_1$ and $\leq_2$ are given according to the following diagram:



$\langle D, \leq_1, \leq_2 \rangle$ is neither interlaced nor distributive. Notice that it is a quite tedious work to check all the conditions, in order to make sure that the depicted bilattice is interlaced. Because D consists out of 9 points, there are 2^9 possible subsets of points. In order to check the interlacing condition, it is necessary to compare all possible pairs of subsets in both order relations. That means we have to perform $2 \cdot 2^9 \cdot 2^9 = 2^{19}$ calculations. In fact, this is a tedious work! To see that $\mathbf{D}$ is not distributive consider the sublattice $\langle \{c, e, g, i, h\}, \leq_1 \rangle$. Consider the following calculation:

$$g \wedge (e \vee h) = g \wedge i = g \neq e = e \vee c = (g \wedge e) \vee (g \wedge h)$$

Therefore, $\mathbf{D}$ cannot be distributive. Moreover, $\mathbf{D}$ is not interlaced, either: Obviously, it holds that $\{c, f, h\} \leq_1 \{f, h\}$. If $\mathbf{D}$ was interlaced, then it would hold: $\Pi\{c, f, h\} \leq_1 \Pi\{f, h\}$. But $g \not\leq_1 h$. Conclude: $\mathbf{D}$ is not interlaced.

(iii) Let $[0, 1] \in \mathbb{R}$ be the closed interval of the real numbers. Consider the Cartesian Product $[0, 1] \times [0, 1]$ with the following order relations $\leq_1$ and $\leq_2$ induced by the standard order relation $\leq$ on the real numbers:

$$\begin{array}{lll} \langle a, b \rangle \leq_1 \langle c, d \rangle & \text{if} & a \leq c \text{ and } b \leq d \\ \langle a, b \rangle \leq_2 \langle c, d \rangle & \text{if} & a \leq c \text{ and } d \leq b \end{array}$$

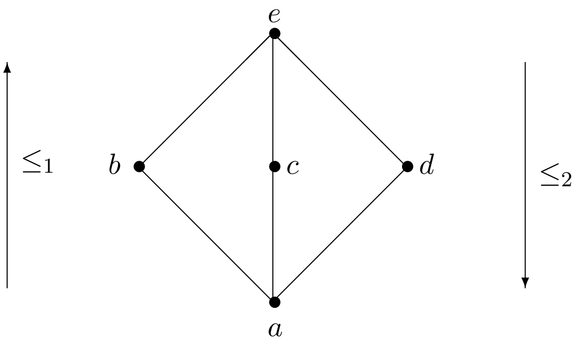
Then it holds: $\langle [0,1] \times [0,1], \leq_1, \leq_2 \rangle$ is an interlaced bilattice. It is easy to check that the interlacing condition is satisfied.

(iv) Assume $[0,1] \in \mathbb{Q}$ is the closed interval of the rational numbers. Consider the pointwise induced order relations as above. Then it turns our that the bilattice $\langle [0,1] \times [0,1], \leq_1, \leq_2 \rangle$ is no longer interlaced, because $\langle [0,1] \times [0,1], \leq_1 \rangle$ as well as $\langle [0,1] \times [0,1], \leq_2 \rangle$ are not complete lattices. For example, the set

$$\{\langle x, y \rangle \in [0,1] \times [0,1] \ \mid \ x = 0 \wedge y \leq 1/\pi\}$$

has no supremum. In [Fi93], these types of bilattices are called (interlaced) pre-bilattices.

Remark 4.2.5 In [Fi89], Melvin Fitting showed that every distributive bilattice is interlaced. The proof is not difficult, therefore we skip it here. It is important to mention that the converse does not hold. Hence, the interlacing condition and the distributive condition are not equivalent. A counterexample is the following structure:



This bilattice is interlaced, because the order relations do not interlace in any interesting way, but not distributive, because both order relations contain the sublattice M_3. Applying the $M_3 - N_5$ theorem of ordinary lattice theory we have the result that this bilattice is not distributive.[14]

Much more examples could be mentioned (and most of them are quite interesting on their own), but we have not the space to do this extensively here. Now, we will come back to Kripke's account of defining a truth predicate in the object language. We have the prerequisites in order to formulate Kripke's ideas precisely in the framework of the theory of interlaced bilattices. The following section is devoted to this construction.

[14] Compare [DaPr90] for further information concerning the $M_3 - N_5$ theorem in ordinary lattice theory.

4.3 The Construction

We presented the basic ideas of Kripke's account in Subsection 4.1.2. What we want to do now is to formulate Kripke's approach in the framework of interlaced bilattices. Using the framework of interlaced bilattices generalizes the original construction in [Kr75]. Although we are working in a non-classical four-valued logic instead of a non-classical three-valued logic (as it was presented in [Kr75]), we can interpret Kripke's original approach as a special case of the general treatment in the theory of interlaced bilattices. In a certain sense, working in the theory of interlaced bilattices enables us to work with a flexible tool in order to model a variety of approaches. To make the ideas precise we consider again the interlaced (and distributive) bilattice FOUR. Recall that $\leq_I$ represents the information order, and $\leq_T$ the truth order. We define the lattice homomorphism $\neg : \{T, F, N, B\} \longrightarrow \{T, F, N, B\}$ as follows: $\neg T = F$, $\neg F = T$, $\neg N = N$, and $\neg B = B$. Notice that the evaluation of all sentences is monotone in the information order (because of the functional monotonicity w.r.t. $\wedge, \vee, \neg$). In the truth order, the evaluation of all positive sentences (i.e. sentences that are not equivalent to a sentence with an odd number of negations in front of it) is monotone (because of the functional monotonicity w.r.t. $+$ and $\cdot$). Notice further that the monotonicity of conjunction and disjunction is a direct consequence of the interlacing property of the underlying bilattice FOUR.

We would like to work in a language L that includes a coding system for arithmetic. That is necessary in order to make sure that Liar-like sentences (or reflexive sentences in general) can be represented in L. Additionally, we assume countably many variables, finitely many constants, moreover terms, atomic formulas, and formulas. Formally, we use the following specifications:

$$
\begin{aligned}
&\Sigma_L = \{c_1, c_2, \ldots, c_n, R_1, R_2, \ldots, R_m, f_1, f_2, \ldots, f_l\}\\
&Var = \{x_1, x_2, \ldots\}\\
&Const = \{c_1, c_2, \ldots, c_n\}\\
&Term = Var \ \cup \ Const \ \cup \ \{f_j(t_1, t_2, \ldots, t_k) \mid t_i \in Var \cup Const \ \wedge \\
&\qquad\qquad f_j \in \{f_1, f_2, \ldots, f_l\}\}\\
&Atom = \{R_i(t_1, t_2, \ldots, t_j) \ \mid \ R_i \text{ is } j\text{-ary} \wedge t_k \in Term\} \cup \\
&\qquad\qquad \{t_n = t_m \mid t_n \in Term \ \wedge \ t_m \in Term\}
\end{aligned}
$$

The set of all formulas $Form$ is defined as the smallest set, such that atomic sentences are in $Form$ and if $\phi \in Form$ and $\psi \in Form$, then $\phi \wedge \psi \in Form$, $\phi \vee \psi \in Form$, and $\neg\psi \in Form$. Furthermore, if $\phi \in Form$, then $\forall x\phi$ is also a formula. Sentences are formulas without free variables.[15] Sentences of a language L are denoted by $Sent_L$.

Assume we add an additional predicate $\mathbf{T}$ to L in order to get $L^+ = L \cup \{\mathbf{T}\}$. We want to find an evaluation function that mirrors Kripke's ideas. We assume that every evaluation function $[[-]] : Sent_L \longrightarrow \{T, F, N, B\}$ satisfies the clauses that are specified in the next definition.

[15] Free variables are inductively defined as usual.

Definition 4.3.1 *Assume we work in a language L and a model $\mathfrak{M} = \langle D, v\rangle$. An evaluation function $[[-]] : Sent_L \longrightarrow \{T, F, N, B\}$ is called classical if $[[-]]$ satisfies the following clauses for every $\phi \in L$.*

$[[\phi]]^{\mathfrak{M}} = v(\phi)$, *if ϕ is atomic*
$[[\phi \wedge \psi]]^{\mathfrak{M}} = [[\phi]]^{\mathfrak{M}} \wedge [[\psi]]^{\mathfrak{M}}$
$[[\neg\phi]]^{\mathfrak{M}} = \neg[[\phi]]^{\mathfrak{M}}$
$[[\forall x\phi]]^{\mathfrak{M}} = \bigwedge_{d\in D}[[\phi(d)]]^{\mathfrak{M}}$
$[[\phi]]^{\mathfrak{M}} = [[\phi]]^{\mathfrak{M}'}$ *for arithmetical ϕ*[16]

In the following, we will work only with classical evaluation functions. Classical evaluation functions make sure that in L everything works like in classical model theory.

It is worth noting that on the left side of the above formulas in Definition 4.3.1 the symbol $\wedge$ is a logical connective in the object language whereas $\wedge$ on the right side is the infimum w.r.t. the order relation $\leq_T$ in the interlaced bilattice FOUR. Although this is quite confusing (and one shouldn't use one and the same symbol for different operations in one formula) it is clear from the context which type of connective we mean. Furthermore, negation $\neg$ is on the left side of the corresponding formula the negation of a formula ϕ and on the right side the lattice homomorphism in FOUR. Again the context makes the reference clear.

Remark 4.3.1 Because FOUR has two order relations and therefore two infima (resp. suprema) operations, it is possible to give a common sense interpretation of the operations in the information order. If one assumes that there are two experts who assign truth values to sentences of a given language, the supremum operation (w.r.t. $\leq_I$) can be interpreted as a certain acceptance operation: every information given by the experts is accepted, even if they are contradictory (as for example in the case $\sup_I\{T, F\} = B$). In this case, every accessible information will be accepted. The infimum operation can be seen as a consensus operation: contradictory information causes the rejection of any judgment ($\inf_I\{T, F\} = N$). In this case, only consistent information is accepted. Although the picture is quite simplified, because every kind of information has equal rights without any further examination concerning the reliability of the source, the model is quite intuitive.

Now we can reformulate the crucial ideas in Kripke's construction. We use the following shorthand: *Arith* denotes the arithmetical sentences. As in the above considerations **T** denotes syntactically the truth predicate.

Definition 4.3.2 *Assume a language $L^+ = L \cup \{\mathbf{T}\}$ and a classical model $\mathfrak{A}$ are given. We call an operator $\Gamma : \{T, F, N, B\}^{Sent_{L^+}} \longrightarrow \{T, F, N, B\}^{Sent_{L^+}}$ a jump operator if for all (possible) valuations $w \in \{T, F, N, B\}^{Sent_{L^+}}$ the*

[16] $\mathfrak{M}'$ is interpreted as the standard model of arithmetic.

following three conditions (i)-(iii) hold:

(i) $\forall \phi \in (L - Arith) : \Gamma(w(\phi)) = [[\phi]]^{\mathfrak{A}}$
(ii) $\forall \phi \in \ Arith : \Gamma(w(\phi)) = [[\phi]]^{\mathfrak{M}}$*, where* $\mathfrak{M}$ *is the standard model of arithmetic*
(iii) $\forall \phi \in Sent_{L^+}$ *of the form* $\phi = \mathbf{T}(\psi) : \Gamma(w(\mathbf{T}(\psi))) = w(\psi)$

The crucial clause (iii) mirrors precisely Tarski's biconditionals. The operator $\Gamma : \{T, F, N, B\}^{Sent_{L^+}} \longrightarrow \{T, F, N, B\}^{Sent_{L^+}}$ does not change the interpretation of arithmetical sentences and sentences that do not contain the truth predicate $\mathbf{T}$. If a sentence ϕ is of the form $\phi = \mathbf{T}(\psi)$, the interpretation of ϕ is reduced to the interpretation of ψ. Notice that Γ is monotone with respect to $\leq_I$. Restricting the range of Γ to positive formulas, Γ is also monotone with respect to $\leq_T$. That is a direct consequence of the interlacing condition of the interlaced bilattice `FOUR`.

We can state the following Lemma that corresponds to well-known lemmas in order theory, the theory of lattices, or the theory of Boolean Algebras. It claims that the function space $\mathbf{D}^X$ induced by an interlaced bilattice $\mathbf{D}$ and an arbitrary set X is also an interlaced bilattice.

Lemma 4.3.3 *Assume* $\boldsymbol{D} = \langle D, \wedge, \vee, \cdot, + \rangle$ *is an interlaced bilattice and assume further that* X *is an arbitrary set. Then, the function space structure* $\mathbf{D}^X = \langle D^X, \leq_1, \leq_2 \rangle$*, such that* $D^X = \{f \mid f : X \longrightarrow D\}$ *and the order relations* $\leq_1$ *and* $\leq_2$ *are induced pointwise,*[17] *is an interlaced bilattice.*

Proof : From order theory it is well-known that if $\langle D, \vee, \wedge \rangle$ is a complete lattice, then $\langle D^X, \vee', \wedge' \rangle$ is a complete lattice as well. The same is true concerning the similar situation with respect to the lattice $\langle D, +, \cdot \rangle$. It remains to show that the resulting structure $D^X = \langle \{f \mid f : X \longrightarrow D\}, \leq_1, \leq_2 \rangle$ preserves the interlacing condition. In order to show this, assume that $F_1 \subseteq \{f \mid X \longrightarrow D\}$ and $F_2 \subseteq \{f \mid f : X \longrightarrow D\}$ are arbitrary subsets of D^X. We have to show that $F_1 \leq_1 F_2 \ \rightarrow \ \Pi F_1 \leq_1 \Pi F_2$ (where Π is the infimum with respect to $\leq_2$). Assume $F_1 \leq_1 F_2$. Then the following equivalences hold by definition:

$$\begin{array}{ll} & F_1 \leq F_2 \\ \Leftrightarrow & (\forall f \in F_1)(\exists g \in F_2) : (f \leq_1 g) \wedge (\forall g \in F_2)(\exists f \in F_1) : (f \leq_1 g) \\ \Leftrightarrow & (\forall f \in F_1)(\exists g \in F_2) : (\forall d \in X : f(d) \leq_1 g(d)) \wedge (\forall g \in F_2)(\exists f \in F_1) : \\ & (\forall d \in X : f(d) \leq_1 g(d)) \end{array}$$

Consider an arbitrary $x \in X$. Then, the following equivalence holds:

$$\begin{array}{ll} & \Pi F_1(x) \leq_1 \Pi F_2(x) \\ \Leftrightarrow & \Pi\{d \in D \mid d = \{f(x) \mid f \in F_1\}\} \leq_1 \Pi\{d \in D \mid d = \{g(x) \mid g \in F_2\}\} \end{array}$$

[17] We define $\leq_1$ and $\leq_2$ as follows: $f_1 \leq_1 f_2$ iff for every $d \in X$ it holds: $f_1(d) \wedge f_2(d) = f_1(d)$ and similarly $f_1 \leq_2 f_2$ iff for every $d \in X$ it holds: $f_1(d) \cdot f_2(d) = f_1(d)$).

The second expression is obviously true, because $\langle D, \wedge, \vee, \cdot, +\rangle$ is an interlaced bilattice: Because $x \in X$ was arbitrarily chosen we can conclude that for every $x \in X$ it holds:

$$\Pi\{d \in D \mid d = \{f(x) \mid f \in F_1\}\} \leq_1 \Pi\{d \in D \mid d = \{g(x) \mid g \in F_2\}\}$$

This is precisely the definition of $\Pi F_1 \leq_1 \Pi F_2$. This suffices to show the lemma.
q.e.d.

Our next step is to state and prove the famous Knaster-Tarski theorem and some further properties of fixed points in complete lattices. This is important, because we need to know the existence of certain fixed points that are candidates for extensions of Kripke's partially defined truth predicates. Additionally, we need to know certain properties of these fixed points.

Lemma 4.3.4 *Assume $\boldsymbol{D} = \langle D, \wedge, \vee\rangle$ is a complete lattice and $\Gamma : D \longrightarrow D$ is a monotone operator. Then, Γ has a maximal and minimal fixed point and the the set of all fixed points* Fix *of Γ is a complete lattice.*

Proof: Consider the set $\text{Up} = \{d \mid d \in D \wedge d \leq \Gamma(d)\}$. Then, the following relation holds: $\forall d \in \text{Up}$: $d \leq \sup(\text{Up})$. Using the property of Γ to be monotone we get for every $d \in \text{Up}$: $\Gamma(d) \leq \Gamma(\sup(\text{Up}))$. Therefore it holds: $\forall d \in \text{Up}$: $d \leq \Gamma(\sup(\text{Up}))$ and because $\sup(\text{Up})$ is the smallest upper bound of Up, it holds: $\sup(\text{Up}) \leq \Gamma(\sup(\text{Up}))$. On the other hand, using monotonicity again we get $\Gamma(\sup(\text{Up})) \leq \Gamma(\Gamma(\sup(\text{Up})))$. That means: $\Gamma(\sup(\text{Up})) \in \text{Up}$. But then: $\Gamma(\sup(\text{Up})) \leq \sup(\text{Up})$. Together: $\sup(\text{Up}) = \Gamma(\sup(\text{Up}))$, or $\sup(\text{Up})$ is itself a fixed point.

Clearly, $\sup(\text{Up})$ is the largest fixed point: Assume d is an arbitrary fixed point, then it holds $d \in \text{Up}$ by the definition of a fixed point, and therefore $d \leq \sup(\text{Up})$. For the existence of the minimal fixed point, consider the bottom element $\perp$ of $\langle D, \leq\rangle$. Clearly, $\perp \leq \Gamma(\perp)$ and because of the monotonicity of Γ we have: $\Gamma(\perp) \leq \Gamma(\Gamma(\perp))$. An easy transfinite induction shows that there is an ordinal λ, such that $\Gamma^\lambda(\perp) = \Gamma^{\lambda+1}(\perp)$, i.e. Γ^λ is a fixed point. Assume $d \in D$ is an arbitrary fixed point. Then, $\perp \leq d$ and because of the monotonicity of Γ, it holds: $\Gamma^\alpha(\perp) \leq \Gamma^\alpha(d)$ for every ordinal α. Therefore, $\Gamma^\lambda(\perp)$ is the smallest fixed point.

It remains to show that the set of all fixed points is a complete lattice. Assume Fix is the set of all fixed points and $X \subseteq \text{Fix}$ is an arbitrary subset of fixed points. Consider the set $Y = \{y \mid y \in \text{Fix} \wedge \sup_D(X) \leq y\}$. Because $\sup(\text{Up}) \in \text{Fix}$, it follows immediately that $Y \neq \emptyset$. The following two conditions hold for Y:

(i) $\sup(X) \leq \inf(Y)$
(ii) for an arbitrary upper bound e of X in Fix we have: $\inf(Y) \leq e$.

Conclude: $\inf(Y)$ is the least upper bound of X in Fix. It remains to show that $\inf(Y)$ is a fixed point. A very similar argument as in the above proof that there is at least one fixed point (Knaster-Tarski Theorem), shows that this is really the case. Together: X has a supremum in Fix. Because the argument for the existence of the infimum is completely similar (and in fact is the dual statement), this suffices to show the lemma. q.e.d.

The following corollary is an easy consequence of the two lemmas above and guarantees the existence of minimal and maximal fixed points in the Kripke account in a four-valued logic (with interlaced bilattices as underlying algebraic structures).

Corollary 4.3.5 *Assume* $\texttt{FOUR} = \langle\{T, F, N, B\}, \leq_I, \leq_T, \neg\rangle$ is given. Assume further that L^+ is a language defined by $L^+ = L \cup \{\mathbf{T}\}$ (where L is given), and $\Gamma : \{T, F, N, B\}^{Sent_{L^+}} \rightarrow \{T, F, N, B\}^{Sent_{L^+}}$ is an operator satisfying the conditions of Definition 4.3.2. Then, there is a maximal and a minimal fixed point in the $\leq_I$ order and the set of fixed points forms a complete lattice (with respect to the $\leq_I$ order).

Proof: This is a trivial consequence of the above lemmas, and the fact that Γ is monotone with respect to $\leq_I$. q.e.d.

We add a further corollary that shows the existence of a minimal and maximal fixed point in both order relations provided we work in positive logic, i.e. there is no negation available in our logic.

Corollary 4.3.6 *Assume the same premises as in Corollary 4.3.5, except that we are working in a logic without negation (positive logic). Then there is a minimal and a maximal fixed point with respect to the* $\leq_I$ *order relation and there is a minimal and maximal fixed point with respect to the* $\leq_T$ *order relation as well.*

Proof: The claim is a consequence of the above Lemmas and the fact that Γ is monotone with respect to $\leq_I$ and $\leq_T$. q.e.d.

The two Corollaries 4.3.5 and 4.3.6 are the basis of Kripke's account to define partial truth predicates. The work that remains to be done is to check whether these fixed points are appropriate for possible extensions of the truth predicate. Criticism and problems of the presented account are postponed to Chapter 6. We add some remarks concerning properties and features of Kripke's construction.

Remark 4.3.2 (i) We chose to work in FOUR. Notice that the negation $\neg$ is assumed to work as described in Example 4.2.3(ii). Clearly, one can work in other interlaced bilattices as well, as long as these bilattices are the algebraic representation of an appropriate logic. Because our aim is to present a framework of a four-valued logic, we chose the special case FOUR as the underlying algebraic structure.

(ii) The fixed points of Corollary 4.3.5 can be associated with the extensions of the possible truth predicates. Similar to Kripke's original work, there are infinitely many fixed points in the present construction.[18] Notice that there is no truth predicate in our construction that corresponds to the maximal intrinsic fixed point of Kripke's original version in a three-valued logic.[19] Every fixed point in Corollary 4.3.5 is intrinsic in Kripke's sense, because interlaced bilattices are essentially two complete lattices connected by a monotonicity condition. Therefore, every fixed point has a supremum and an infimum with any other fixed point. That makes it difficult to designate a special fixed point as the intuitively correct extension of the truth predicate. On the other hand, there are crucial reasons to doubt whether the maximal intrinsic fixed point is really a good choice for the truth predicate. In Chapter 6, we will consider various reasons, why this is the case.

(iii) One can extract Kripke's construction as a special case of the present account. The only thing that we need to do is to restrict the evaluation to sentences that are either true, false, or neither true nor false prior to the application of the monotone operator to the set $\{T, F, N, B\}^{Sent_{L^+}}$. A dual move is possible by restricting the evaluation to sentences that are either true, false, or both true and false. The resulting construction has certain similarities to paraconsistent logic, a logic that was proposed by Priest.[20]

(iv) Concerning the Truth-teller sentence, we have the following behavior: in the maximal fixed point, the Truth-teller sentence is both true and false with respect to the $\leq_I$ order, whereas this sentence is neither true nor false in the minimal fixed point with respect to the $\leq_I$ order. The Liar sentence has a similar behavior: in the maximal fixed point, this sentence is true and false, whereas in the minimal fixed point this sentence is neither true nor false. When switching to positive logic as described in Corollary 4.3.6, we have the following situation: in the maximal fixed point with respect to the $\leq_T$ order relation the truth value of the Truth-teller sentence is T, whereas in the minimal fixed point, this sentence has the truth value F.

(v) A last remark concerns the monotone operator Γ. From an abstract point of view, every monotone operator Γ would give us what we want, namely fixed points. In general, an arbitrary monotone operator would not give us a reasonable extension of the truth predicate. By choosing the conditions of Definition 4.3.2 we model the intuition behind a truth predicate, especially we can preserve Tarski's biconditionals.

[18] For further information concerning the number of fixed points in different kinds of three-valued logics compare [CaDa91]. For four-valued logics a quite similar situation holds.

[19] For more information concerning the maximal fixed intrinsic point in a three-valued logic, the reader is referred to Corollary 4.4.3.

[20] Cf. [Pr79].

In the following section, we will compare the presented construction with Kripke's original construction in a three-valued logic. Whereas the presented four-valued approach has the advantage to be quite straightforward working in three-valued logic is slightly more complicated.

4.4 Discussion

In this section, we will present the original construction of Kripke in a three-valued logic. The general idea remains the same as described in Section 4.3, except that the underlying logic has different algebraic properties. Order theoretically, Kripke developed his account in a semilattice. Later Visser showed in [Vi89] that the order theoretic structure that corresponds nicely to the strong Kleene-tables are CCPOs. In Definition 4.2.1(ii), we specified the concept 'CCPO'. In order to make clear the differences between the three-valued case and the four-valued case, we state the crucial theorems that are needed to develop Kripke's construction in a CCPO. The result will be that the construction and the corresponding theorems are slightly more complicated and not as straightforward as in the case of the four-valued case. Before we can do this, we need to make precise what it means for a point to be intrinsic.[21]

Definition 4.4.1 *Assume $\langle D, \leq \rangle$ is a CCPO. A point $d \in D$ is called intrinsic in D if and only if for all $d' \in D$ it holds:* $\sup\{d, d'\}$ *exists.*

Intrinsic points are points that are consistent with all other points (or with a certain subset of other points in the case we restrict those appropriately). From an information theoretic perspective one can associate with an intrinsic point an object that is consistent with every information theoretic extension. For example, if we are working in a CCPO representing three truth values $\{T, F, N\}$, then N is consistent with T and F, because a statement that is underspecified can be extended to a true statement as well as to a false statement without creating inconsistencies.

The following theorem states some properties of monotone operators defined on CCPOs.

Theorem 4.4.2 *Assume $\mathbf{D} = \langle D, \leq \rangle$ is a CCPO, and $\Gamma : D \longrightarrow D$ is a monotone operator. Then the following holds:*

(i) The set of fixed points of Γ is not empty. Furthermore, the set of fixed points forms again a CCPO.
(ii) The set of intrinsic fixed points S forms a complete lattice.
(iii) If S' denotes the set of all intrinsic points in D, then it holds:

$$\sup(S') = \inf\{d \in D \mid d \text{ is maximal}\}$$

[21] The concept of intrinsic points goes back to [Kr75].

Proof: (i) Let $\perp$ be the bottom element of $\mathbf{D}$. Because of the monotonicity of Γ it holds obviously (for α to be an arbitrary ordinal):

$$\perp \leq \Gamma(\perp) \leq \Gamma(\Gamma(\perp)) \leq ... \leq \Gamma^{\alpha}(\perp) \leq \dots$$

Because a CCPO is chain-complete it follows that

$$\sup\{\perp, \Gamma(\perp), \Gamma(\Gamma(\perp)), \dots, \Gamma^{\alpha}(\perp), \dots\}$$

exists. Moreover, this supremum is an element in the chain itself (again because of chain-completeness). Therefore we get:

$$\Gamma(\sup\{\perp, \Gamma(\perp), \Gamma(\Gamma(\perp)) \dots\}) \leq \sup\{\perp, \Gamma(\perp), \Gamma(\Gamma(\perp)) \dots\}$$

On the other hand, because of the monotonicity of Γ we can also establish:

$$\sup\{\perp, \Gamma(\perp), \Gamma(\Gamma(\perp)) \dots\} \leq \Gamma(\sup\{\perp, \Gamma(\perp), \Gamma(\Gamma(\perp)) \dots\})$$

Together we can conclude: $\sup\{\perp, \Gamma(\perp), \Gamma(\Gamma(\perp)) \dots\}$ is a fixed point.

In order to show that the set of fixed points Fix forms again a CCPO, notice that the constructed fixed point above is the minimal fixed point (therefore $\sup\{\perp, \Gamma(\perp), \Gamma(\Gamma(\perp)) \dots\}$ is the bottom element of the set of fixed points). It is clear that the set of fixed points forms a partially ordered set. To check the coherence condition for CCPOs is straightforward.

(ii) Let $S = \{x \mid x \in \text{Fix } \wedge x \text{ intrinsic}\}$ be the set of intrinsic points. We have to show that every subset $X \subseteq S$ has an infimum and a supremum. Clearly, $\langle S, \leq \rangle$ is a partial order. First, we need to show that $\sup(S)$ exists and is itself intrinsic. Notice that S is consistent. Therefore $\sup(S)$ exists. We show that $\sup(S)$ is itself in S. Because $\sup\{s, d\}$ exists for all $d \in D$ and $s \in S$, it follows that $\{S, d\}$ is consistent. Therefore, $\sup\{S, d\}$ exists. The following inequalities hold (trivially):

$\sup(S) \leq \sup\{S, d\}$
$d \leq \sup\{S, d\}$

Hence, $\{\sup(S), d\}$ is consistent and $\sup\{\sup(S), d\}$ exists, because we are working in a CCPO. Because this holds for every $d \in D$, $\sup(S)$ is itself intrinsic.

With these prerequisites we can prove claim (ii). Assume $X \subseteq S$ is arbitrarily chosen. It is clear that $\inf(X)$ exists and $\inf(X) \in D$. Because $S \subseteq \text{Fix}_{\Gamma}$ (Fix_{Γ} denotes the set of fixed points of Γ) it holds: $\inf(X) \in \text{Fix}_{\Gamma}$. We need to show that $\inf(X)$ is itself intrinsic. Because $\inf(X) \leq \sup(S)$, for all $d \in D$ we have the following relations:

$\inf(X) \leq \sup\{S, d\}$
$d \leq \sup\{S, d\}$

This means: $\inf(X)$ is consistent and therefore, $\sup\{\inf(X), d\}$ exists (because of the properties of a CCPO). Because $d \in D$ was arbitrarily chosen, we can conclude that $\inf(X)$ is intrinsic.

A similar consideration yields the corresponding claim concerning the supremum of X for $X \subseteq S$. Now, it holds:

$$\forall x, y \in X : x \leq \sup(S) \wedge y \leq \sup(S)$$

From that fact it follows that X is consistent and therefore $\sup(X)$ exists. Now the following inequalities hold:

$\sup(X) \leq \sup\{S, d\}$
$d \leq \sup\{S, d\}$

As above we can conclude: $\sup\{\sup(X), d\}$ exists for all $d \in D$ and therefore we have the desired result $\sup(S) \in S$.

(iii) Let S' be the set of all intrinsic points in D. In order to show that it holds:

$\sup(S') = \inf\{x \mid x \in D \wedge x \text{ is maximal}\}$

one has to show two facts. First, we need to show that $\inf\{x \mid x \in D \ \wedge \ x$ is maximal$\}$ is intrinsic and second, we need to show that $\sup(S')$ is a lower bound of $\{x \mid x \in D \ \wedge \ x$ is maximal$\}$. Notice that $\inf\{x \mid x \in D \ \wedge \ x$ is maximal$\}$ exists. Let $d \in D$ be arbitrarily chosen. Then, there exists $y \in \{x \mid x \in D \ \wedge \ x$ is maximal$\}$, such that it holds:

$d \leq y$
$\inf\{x \mid x \in D \ \wedge \ x \text{ is maximal}\} \leq y$

Then, the set $\{d, \inf\{x \mid x \in D \ \wedge \ x$ is maximal$\}\}$ is consistent and therefore $\sup\{d, \inf\{x \mid x \in D \ \wedge \ x$ is maximal$\}\}$ exists. Because $d \in D$ was arbitrarily chosen, conclude that $\inf\{x \mid x \in D \ \wedge \ x$ is maximal$\}$ is intrinsic. Let $y' \in \{x \mid x \in D \ \wedge \ x$ is maximal$\}$ be arbitrarily chosen. Then, we have: $\sup\{\sup(S'), y'\}$ exists. Because y' is maximal, it holds: $\sup(S') \leq y'$. Conclude: $\sup(S')$ is a lower bound of $\{x \mid x \in D \ \wedge \ x$ is maximal$\}$. Together we have shown that it holds: $\sup(S') = \inf\{x \mid x \in D \ \wedge \ x$ is maximal$\}$ q.e.d.

The proof for claim (i) of the theorem (namely that there is at least one fixed point and that the fixed point structure forms a CCPO) is slightly more complicated than the famous Knaster-Tarski Theorem for lattices. In a certain sense, (i) is a generalization of the classical Knaster-Tarski Theorem. Instead of considering lattices, we consider weaker structures, namely CCPOs. It should be mentioned that a similar theorem to Theorem 4.4.2 can be proven even for

CPOs.[22]

We can state an easy but very important corollary based on the above theorem. This corollary claims the existence of a maximal intrinsic fixed point. For Kripke's account the existence of this designated point is crucial because the maximal intrinsic fixed point gives Kripke a possible extension of the desired truth predicate.

Corollary 4.4.3 *Assume $\langle D, \leq \rangle$ is a CCPO, and $\Gamma : D \longrightarrow D$ is a monotone operator. Then it holds: there exists a maximal intrinsic fixed point.*

Proof: This is a trivial consequence of Theorem 4.4.2(ii). q.e.d.

The proof of the existence of the maximal intrinsic fixed point does not contain transfinite ordinals at any point. This is different, if one works in the framework of CPOs. Although even in CPOs the existence of the maximal intrinsic fixed point is provable, in the standard proof one needs a transfinite induction for this result. It is an open question, whether there exists a proof that does not use ordinals.[23]

We state the last fact that is important to establish Kripke's account in three-valued logic. This is the following fact: assume a CCPO $\mathbf{D} = \langle D, \leq \rangle$ is given. Then the function space $\mathbf{D}^X$ where X is an arbitrary set is again a CCPO.

Fact 4.4.4 *Assume $\mathbf{D} = \langle D, \leq \rangle$ is a CCPO. Then, the function space $\mathbf{D}^X = \langle D^X, \leq' \rangle$ is also a CCPO.*

Proof: Apply a similar argument as in the corresponding proofs for bilattices, Boolean algebras, lattices etc. q.e.d.

Corollary 4.4.5 *Assume $L^+ = L \cup \{\mathbf{T}\}$ is an extension of a given language L. Assume further that $\Gamma : \{T, F, N\}^{Sent_{L^+}} \longrightarrow \{T, F, N\}^{Sent_{L^+}}$ is an operator satisfying the conditions of Definition 4.3.2. Then there exists a maximal intrinsic fixed point of Γ.*

Proof: The claim is an easy consequence of Corollary 4.4.3, Fact 4.4.4, and the fact that $\langle \{T, F, N\}, \leq \rangle$ is a CCPO.[24] q.e.d.

Remark 4.4.1 (i) The above theorem holds also for weaker structures like CPOs and complete semilattices. In [Ku96a], there is a detailed discussion of this topic. In this work, it is shown that the complete considerations can

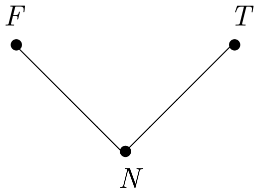
[22] Cf. [Ku96a].

[23] Cf. [Ku96a].

[24] For further information concerning the structure $\langle \{T, F, N\}, \leq \rangle$ compare Remark 4.4.1(ii).

be performed in CPOs, semilattices, as well as CCPOs. In a certain sense, the interpretation of the underlying algebraic structure as CCPO, CPO, or semilattice etc. is up to the one who uses it. There are certain reasons that support the claim that the natural algebraic structure for Kleene's three-valued logic are CCPOs, because CCPOs are the strongest structures that are preserved in $\langle \{T, F, N\}^{Sent_{L^+}}, \leq \rangle$.[25] Nevertheless, the properties of the modeling of Kripke's ideas using three-valued logic remains quite similar independently of the usage of CCPOs, CPOs, or complete semilattices.

(ii) Notice that the translation of these order theoretic results into a Kripke-style approach can be done by considering the three valued logic $\langle \{T, F, N\}, \leq \rangle$ where the order relation is given by the following diagram in which T represents *true*, F represents *false*, and N represents *neither true nor false*:



The above structure is a CCPO (also a CPO, as well as a complete semilattice). Notice further that for a given language L^+ the structure $\langle \{T, F, N\}^{Sent_{L^+}}, \leq \rangle$ where $\leq$ is induced pointwise is a CCPO, too.

Whereas in Kripke's original account, intrinsic fixed points where motivated via their importance for the resulting truth theory, this is no longer true if one works in the framework of interlaced bilattices. In Section 4.3, we saw that every fixed point is intrinsic and the set of intrinsic fixed points forms a complete lattice. Therefore, it is not of any interest to ask, whether there is a maximal intrinsic fixed point. Trivially, there is a maximal intrinsic fixed point, because every fixed point is intrinsic. Although even in a three-valued logic the set of intrinsic fixed points forms a complete lattice, there is quite a lot of work to do in order to prove the desired result. In four-valued logic, everything is much easier in character, because the underlying algebraic structure requires a lower amount of work.

A general discussion whether the maximal intrinsic fixed point is a reasonable and appropriate choice for the truth predicate for natural languages as well as a discussion of other problematic features of Kripke's account will be postponed to Chapter 6.

[25] Cf. [Vi89].

In three-valued logic, it is clear that the Liar sentence will have the truth value N in the maximal intrinsic fixed point. The same is true for the Truth-teller: the Truth-teller must be neither true nor false in the maximal intrinsic fixed point. The account makes also clear that there is a maximal fixed point that makes the truth-teller false and another maximal fixed point that makes the Truth-teller true. The Liar sentence remains neither true nor false in every maximal fixed point.

One can interpret the four-valued approach in the framework of interlaced bilattices as a generalization of Kripke's original work. This is supported by the fact that Kripke's account is a special case of the four-valued account by restricting the possible truth values of atomic formulas to the truth values $\{T, F, N\}$. An evaluation of complex sentences according to the rules of the truth order is equivalent to the original Kripke account. But on the other hand the four-valued approach is stronger: as we can see later (compare Section 5.5), we can detect relations between different interpretations of pathological sentences that are not possible in Kripke's original account.

Kripke argued for the preference of the maximal intrinsic fixed point in order to get a good interpretation of the truth predicate of natural language. One reason for the choice of this special fixed point is the fact that the maximal intrinsic fixed point contains more information than the minimal fixed point, and on the other hand this point is consistent with every other fixed point (because it is intrinsic), i.e. does not contain too much information. It is a highly controversial question, whether Kripke's arguments concerning this point are convincing or not.

Because of the fact that the corresponding algebraic structure of four-valued logic (as an extension of the strong Kleene-tables) is a complete lattice, other interpretations of the underlying algebraic structure are possible. Another account is to model the inductively defined truth predicate in the ordinary truth order of classical four-valued logic. The standard reference for this account is [Vi84]. As in the classical Kripke case, it is possible to distinguish paradoxical sentences, grounded sentences, and biconsistent sentences. A sentence ϕ is called grounded false, if the sentence has truth value F in every fixed point. ϕ is called biconsistent, if there are fixed points where the sentence is false and there are fixed points where the sentence is true. A sentence ϕ is called paradoxical, if ϕ has truth value N or B in every fixed point.

We finish this chapter with these remarks. In the next chapter, we will examine properties of different classes of interlaced bilattices more closely, in particular we will prove certain characterization theorems for interlaced bilattices.

4.5 History

Mathematically, the presented account goes back to [Mo74] in which the theory of inductive definitions were developed. Kripke used essentially Moschovakis' ideas in order to get an inductively defined partial truth definition. Furthermore, as predecessors of theories of truth before 1975 we mention Fitch, Scott, Aczel, and Feferman. These authors developed a theory of truth that is formulated in classical logic by restricting other aspects of the theory (like unbounded quantification or the definable functions). The original field in which these theories were developed was combinatory logic and the λ-calculus. References are [Sc75, Fe84], and [Ac78].

The described construction to define partial truth predicates was first developed by Kripke and Martin and Woodruff (in [Kr75] and [MaWo75]). Whereas both deserve equal credit for their constructions of partially defined truth predicates, usually Saul Kripke is credited for this progress. The existence of the maximal intrinisic fixed point in semilattices was claimed by Kripke (Cf. [Kr75]) and proven for CCPOs (in a quite different context) in [MaSh76]). Other important theorems are further developments of work known from the theory of partial orders and lattice theory. A good overview of the topic and the possible algebraic structures can be found in [Vi89] and [Ku96a].

The origin of the theory of bilattices was the area of theoretical computer science and artificial intelligence. Whereas Ginsberg (in [Gi86, Gi88]) worked with bilattices and with distributive bilattices, it was Melvin Fitting (in [Fi88, Fi89, Fi91, Fi93, Fi94]) who discovered that interlaced bilattices are the structures needed in order to model several topics in theoretical computer science and in particular, Kripke's fixed point approach. The construction as developed in section 4.3 is due to Fitting.

Criticism concerning Kripke's ideas was mentioned in an enormous number of papers and books. It would be impossible to mention all these authors. Good overviews of criticism are given in [Mc91, GuBe93, Ma84] and in [BarEt87] (to mention only some important ones).

Chapter 5

Properties of Bilattices

In this chapter, we will prove three characterization theorems for interlaced bilattices and certain subclasses of interlaced bilattices. The characterization theorems have two main advantages. First, they provide a better understanding of the properties for the considered classes of bilattices. Second, they can be used to show further properties of Kripke's fixed point approach. In this chapter, we will restrict our attention mainly to the technical details of the constructions. Some applications are mentioned in Section 5.5.

5.1 A Characterization of Interlaced Bilattices

We begin this section with recalling some definitions that will crucially be used in this chapter.

Definition 5.1.1 *(i) For arbitrary and nonempty subsets A and B of a bilattice $\langle D, \leq_1, \leq_2 \rangle$ the epression $A \leq_1 B$ denotes the following relation:*

$$(\forall a \in A)(\exists b \in B) : a \leq_1 b \land (\forall b \in B)(\exists a \in A) : a \leq_1 b$$

The expression $A \leq_2 B$ is defined similarly.
(ii) For arbitrary elements $x, y \in D$, the expression $x <_1 y$ means that $x \leq_1 y$ and $x \neq y$. Concerning $\leq_2$ a similar relation is defined. The expression $A \leq_i B$ for $A, B \subseteq D$ and $i \in \{1, 2\}$ is analogously defined.

The following definition gives us a construction method how to generate bilattices using ordinary lattices. In a certain sense, the next definition specifies the product of two lattices, such that the induced order relations on the product preserve certain features, namely the features of an interlaced bilattice. Notice that the following definition is a generalization of Definition 4.2.4 extending chains to complete lattices.

Definition 5.1.2 *Let $\langle D, \leq \rangle$ be a complete lattice. Then the structure $D \star D = \langle D \times D, \leq_1, \leq_2 \rangle$ is defined as follows:*

$$(x_1, x_2) \leq_1 (y_1, y_2) \;\Leftrightarrow\; x_1 \leq y_1 \land y_2 \leq x_2$$

$$(x_1, x_2) \leq_2 (y_1, y_2) \Leftrightarrow x_1 \leq y_1 \wedge x_2 \leq y_2$$

There are a number of results concerning properties of and relations between certain types of bilattices. Without trying to be complete we will mention some of them.

Fact 5.1.3 *(i) Every distributive bilattice is interlaced.*
(ii) If $\langle D, \leq_1, \leq_2 \rangle$ is an interlaced bilattice, then the following four equations hold:

(α) $\Pi D \wedge \Sigma D = \bigwedge D$
(β) $\Pi D \vee \Sigma D = \bigvee D$
(γ) $\bigwedge D \cdot \bigvee D = \Pi D$
(δ) $\bigwedge D + \bigvee D = \Sigma D$

(iii) Let $\langle D, \leq \rangle$ be a complete lattice. Then $D \star D$ is an interlaced bilattice.
(iv) Let $\langle D, \leq \rangle$ be a complete and distributive lattice. Then $D \star D$ is a distributive bilattice.
(v) Let $\langle \{0, 1, 2, ..., n\}, \leq \rangle$ and $\langle \{0, 1, 2, ..., m\}, \leq \rangle$ be two chains where $\leq$ is the standard order relation on the natural numbers $\mathbb{N}$. Then the product bilattice $\mathbf{D}$ specified by $\mathbf{D} = \langle \{0, 1, 2, ..., n\} \times \{0, 1, 2, ..., m\}, \leq_1, \leq_2 \rangle$ is an interlaced bilattice.

Proof: (i) Compare [Fi89].
(ii) Compare [Fi91].
(iii) Compare [Fi89].
(iv) Compare [Gi88].
(v) This is simply a special case of claim (iv). q.e.d.

We add some remarks concerning the claims above that are immediate consequences of the properties of bilattices.

Remark 5.1.1 (i) The reverse direction of Fact 5.1.3(i) is not true in general. In Section 5.3 we will specify a sufficient condition to guarantee that the reverse direction is also true (i.e. that an appropriately restricted class of bilattices determines a class of distributed bilattices).
(ii) It is clear that the product bilattice of complete chains is also a distributive bilattice. This follows immediately from Fact 5.1.3(iv) and the obvious fact that every chain is distributive.
(iii) Fact 5.1.3(ii) is a special case of the general Theorem 5.1.10 we will prove in this section later.

In the following sections, we will restrict our considerations to properties of interlaced bilattices and to properties of certain subclasses of interlaced bilattices. First, we will show that the interlacing condition is a relatively strong restriction of bilattices. To put it in a (not very precise) statement: every interlaced bilattice is 'rhombus-like' and every 'rhombus-like' bilattice is interlaced.

In order to get a characterization of interlaced bilattices, we formulate some lemmas which give us necessary conditions for bilattices to be interlaced. The first lemma claims that the infimum w.r.t. $\leq_i$ of every subset of an interlaced bilattice is smaller w.r.t. $\leq_j$ than the supremum of this subset w.r.t. $\leq_j$, and bigger w.r.t. $\leq_j$ than the infimum of this subset w.r.t. $\leq_i$. A similar condition holds for the supremum of an arbitrary subset. Formally, we can state this in the following lemma.

Lemma 5.1.4 *If $\langle D, \leq_1, \leq_2 \rangle$ is an interlaced bilattice, then the following conditions hold:*

(i) for all $A \subseteq D$ with $A \neq \emptyset : \bigwedge A \leq_1 \Pi A \leq_1 \bigvee A$ and $\bigwedge A \leq_1 \Sigma A \leq_1 \bigvee A$
(ii) for all $A \subseteq D$ with $A \neq \emptyset : \Pi A \leq_2 \bigwedge A \leq_2 \Sigma A$ and $\Pi A \leq_2 \bigvee A \leq_2 \Sigma A$
(iii) for all $A \subseteq D$ with $A \neq \emptyset : \bigwedge A \leq_1 \Pi A \wedge \Sigma A \leq_1 \Pi A \vee \Sigma A \leq_1 \bigvee A$
(iv) for all $A \subseteq D$ with $A \neq \emptyset : \Pi A \leq_2 \bigwedge A \cdot \bigvee A \leq_2 \bigwedge A + \bigvee A \leq_2 \Sigma A$

Proof: (i) Let $A \subseteq D$ be an arbitrary set. Then, the relation $\bigwedge A \leq_1 a$ holds for every $a \in A$. Hence, it holds by definition: $\bigwedge A \leq_1 A$. Because of the interlacing condition of $\langle D, \leq_1, \leq_2 \rangle$, we can deduce: $\Pi(\bigwedge A) = \bigwedge A \leq_1 \Pi A$ and with the same reasoning we get the relation $\Sigma(\bigwedge A) = \bigwedge A \leq_1 \Sigma A$.
On the other hand, let again $A \subseteq D$ be an arbitrary set. Then: $A \leq_1 \bigvee A$. Using again the interlacing condition of $\langle D, \leq_1, \leq_2 \rangle$ we get: $\Pi A \leq_1 \Pi(\bigvee A) = \bigvee A$ and $\Sigma A \leq_1 \Sigma(\bigvee A) = \bigvee A$. Together we can deduce for nonempty arbitrary sets $A \subseteq D$ the inequalities: $\bigwedge A \leq_1 \Pi A \leq_1 \bigvee A$ and $\bigwedge A \leq_1 \Sigma A \leq_1 \bigvee A$.

(ii) The assertion for the $\leq_2$ order relation is the dual statement of (i). Hence, it is a direct consequence of claim (i).

(iii)/(iv) The assertions are an immediate consequence of (i) and (ii), with reasoning in ordinary lattice theory. q.e.d.

The next Lemma shows that in an interlaced bilattice there is a closer connection between the suprema and infima of arbitrary sets (with respect to their order-relations). We are able to formulate necessary conditions for the interlacing property of a given bilattice. As is intuitively clear from the definition of the interlacing condition, we need a connection between the order relations $\leq_1$ and $\leq_2$. The following Lemma clarifies this connection. Moreover, the next Lemma shows one direction of the characterization theorem 5.1.10.

Lemma 5.1.5 *Let $\mathbf{D} = \langle D, \leq_1, \leq_2 \rangle$ be an interlaced bilattice. Then the following relations hold:*

(i) for all nonempty $A \subseteq D : \bigwedge A = \Pi A \wedge \Sigma A$ and
for all nonempty $A \subseteq D : \Pi A = \bigwedge A \cdot \bigvee A$

(ii) for all nonempty $A \subseteq D : \bigvee A = \Pi A \vee \Sigma A$ and
for all nonempty $A \subseteq D : \Sigma A = \bigwedge A + \bigvee A$

Proof: (i) Let $A \subseteq D$ be an arbitrary set ($A \neq \emptyset$). Because of the relations $\Pi A \leq_2 \bigwedge A$ and $\Sigma A \leq_2 \Sigma A$ (the first claim is justified by 5.1.4) it follows using the interlacing property of $\langle D, \leq_1, \leq_2 \rangle$: $\Pi A \wedge \Sigma A \leq_2 \bigwedge A \wedge \Sigma A$. Using again 5.1.4 we get $\bigwedge A \wedge \Sigma A = \bigwedge A$. Hence, we can deduce the relation $\Pi A \wedge \Sigma A \leq_2 \bigwedge A$. Concerning the other direction, we argue as follows: it holds obviously $\bigwedge A \leq_2 \Sigma A$ and $\Pi A \leq_2 \Pi A$ (again the first claim is justified by 5.1.4). With the interlacing condition we can deduce the relations $\bigwedge A = \bigwedge A \wedge \Pi A \leq_2 \Pi A \wedge \Sigma A$. Conclude: $\Pi A \wedge \Sigma A = \bigwedge A$.
For the second part of claim (i) let again $A \subseteq D$ be an arbitrary, nonempty set. A similar reasoning as above establishes the relations: $\bigwedge A \leq_1 \Pi A$ and $\bigvee A \leq_1 \bigvee A$. Therefore, we have: $\bigwedge A \cdot \bigvee A \leq_1 \Pi A \cdot \bigvee A = \Pi A$ (by the interlacing condition of $\langle D, \leq_1, \leq_2 \rangle$ and 5.1.4). The other direction, namely that $\Pi A = \Pi A \cdot \bigwedge A \leq_1 \bigwedge A \cdot \bigvee A$, is justified by the facts that $\Pi A \leq_1 \bigvee A$, that $\bigwedge A \leq_1 \bigwedge A$, and that $\langle D, \leq_1, \leq_2 \rangle$ is interlaced. Together we get the relation $\Pi A = \bigwedge A \cdot \bigvee A$.

(ii) We apply a similar reasoning as in (i): For example, we have: $\Pi A \leq_2 \Pi A$ and $\bigvee A \leq_2 \Sigma A$ and therefore (using again the interlacing condition) $\bigvee A = \Pi A \vee \bigvee A \leq_2 \Pi A \vee \Sigma A$. The other direction is justified because of the following reasoning: $\Pi A \leq_2 \bigvee A$ and $\Sigma A \leq_2 \Sigma A$ implies $\Pi A \vee \Sigma A \leq_2 \bigvee A \vee \Sigma A = \bigvee A$. Together we have: $\bigvee A = \Pi A \vee \Sigma A$. q.e.d.

The above Lemma 5.1.5 determines necessary conditions for a bilattice to be interlaced. In order to get a characterization result, we need to prove that these conditions are also sufficient. Unfortunately, we cannot do this directly. We take a detour by stating sufficient conditions that seemingly are very specific properties of a bilattice. These properties will be generalized to the interlacing condition later.

Lemma 5.1.6 *Let $\langle D, \leq_1, \leq_2 \rangle$ be a bilattice, i.e. $\langle D, \leq_1 \rangle$ and $\langle D, \leq_2 \rangle$ are complete lattices, such that $\langle D, \leq_1, \leq_2 \rangle$ satisfies the following properties:*

For all nonempty $A \subseteq D : \bigwedge A = \Pi A \wedge \Sigma A$ and
for all nonempty $A \subseteq D : \Pi A = \bigwedge A \cdot \bigvee A$ and
for all nonempty $A \subseteq D : \bigvee A = \Pi A \vee \Sigma A$ and
for all nonempty $A \subseteq D : \Sigma A = \bigwedge A + \bigvee A$

Then it holds: If $A \subseteq D$ and $B \subseteq D$ are arbitrary nonempty sets and $E \subseteq \{e \mid \forall a \in A : a \leq_1 e\}$ is a nonempty set such that the following properties hold: $B = A \cup E \cup \bigvee E$, and $\bigvee A <_1 \bigvee B$, then conditions (i) and (ii) hold:
(i) $\Pi A \leq_1 \Pi B$
(ii) $\Sigma A \leq_1 \Sigma B$

Dually, if $A, B \subseteq D$ are arbitrary nonempty sets and $\emptyset \neq E \subseteq \{e \mid \forall a \in A : a \leq_2 e\}$, such that the following properties hold: $B = A \cup E \cup \Sigma E$ and $\Sigma A <_2 \Sigma B$, then conditions (iii) and (iv) hold:

(iii) $\bigwedge A \leq_2 \bigwedge B$
(iv) $\bigvee A \leq_2 \bigvee B$

Proof: (i) Let $A \subseteq D$ be a nonempty set, such that $B = A \cup E \cup \bigvee E$, $\bigvee A <_1 \bigvee B$, $E \subseteq \{e \mid \forall a \in A : a \leq_1 e\}$, and $C = \bigvee A \cup E \cup \bigvee E$. Then it holds: $\Pi C = \bigwedge C \cdot \bigvee C = \bigvee A \cdot \bigvee C$ and $\bigvee A \leq_1 \Pi C$. Consider $x := \Pi A \cdot \Pi C$. Because of the relation $x \leq_2 \Pi A$ it holds for all $a \in A$: $x \leq_2 a$ and similarly because of the relation $x \leq_2 \Pi C$, it is justified that for all $c \in C : x \leq_2 c$. Hence, we can conclude for $B = A \cup C$ that the following relation holds: $x \leq_2 \{b \mid b \in B\}$. To prove that x is the greatest lower bound of B with respect to $\leq_2$, let y be an arbitrary lower bound of B with respect to $\leq_2$. Then it is evident that $y \leq_2 \Pi C$ and $y \leq_2 \Pi A$ hold. Because $x = \Pi A \cdot \Pi C$ we can deduce the relation $y \leq_2 x$. We can conclude: $x = \Pi A \cdot \Pi C = \Pi B$. Because of the inequality $\Pi A = \Pi A \wedge \Pi C \leq_1 \Pi A \cdot \Pi C = \Pi B$ (justified by the premises and the fact that $\Pi A \leq_1 \bigvee A \leq_1 \Pi C$), we can deduce the claim, namely that it holds $\Pi A \leq_1 \Pi B$.

(ii) The proof for (ii) is quite similar to the proof of claim (i). Let $A \subseteq D$ be an arbitrary set, $B = A \cup E \cup \bigvee E$, such that $\bigvee A <_2 \bigvee B$, $E = \{e \mid \forall a \in A : a \leq_1 e\}$, and $C = \bigvee A \cup E \cup \bigvee E$. Then $\Sigma C = \bigwedge C + \bigvee C = \bigvee A + \bigvee C$ and $\bigvee A \leq_1 \Sigma A$. Consider now $x := \Sigma A + \Sigma C$. Because of the relation $\Sigma A \leq_2 x$, it holds $\forall a \in A : a \leq_2 x$ and similarly because of the relation $\Sigma C \leq_2 x$, it is justified that $\forall c \in C : c \leq_2 x$. Hence, we can conclude for $B = A \cup C$ that the relation $B \leq_2 x$ holds. To prove that x is the lowest upper bound of B with respect to $\leq_2$, let y be an arbitrary upper bound of B with respect to $\leq_2$. Then it is evident that $\Sigma C \leq_2 y$ and $\Sigma A \leq_2 y$, and because $x = \Sigma A + \Sigma C$, the relation $x \leq_2 y$ holds. We can conclude: $x = \Sigma A + \Sigma C = \Sigma B$. Because of the inequality $\Sigma A = \Sigma A \wedge \Sigma C \leq_1 \Sigma A + \Sigma C = \Sigma B$ (essentially justified by the fact that $\Sigma A \leq_1 \bigvee A \leq_1 \Sigma C$), we can deduce the desired assertion $\Sigma A \leq_1 \Sigma B$.

(iii) This claim is the dual statement of (i).

(iv) This claim is the dual statement of (ii). q.e.d.

Now, we will formulate the dual assertion of Lemma 5.1.6. We claim the following: if a subset A of a bilattice is a proper subset of a set B (of the same bilattice) and B has only additional elements which are smaller than every $a \in A$, then (assuming certain additional conditions) the sets A and B satisfy the interlacing property.

Lemma 5.1.7 *Let $\langle D, \leq_1, \leq_2 \rangle$ be a bilattice, such that $\langle D, \leq_1, \leq_2 \rangle$ satisfies the following properties:*

For all nonempty $A \subseteq D : \bigwedge A = \Pi A \wedge \Sigma A$ and
for all nonempty $A \subseteq D : \Pi A = \bigwedge A \cdot \bigvee A$ and
for all nonempty $A \subseteq D : \bigvee A = \Pi A \vee \Sigma A$ and

for all nonempty $A \subseteq D : \Sigma A = \bigwedge A + \bigvee A$

Then it holds: If $B \subseteq D$ *is an arbitrary set and* $A = B \cup E \cup \bigwedge E$ *for a nonempty set* $E = \{e \mid \forall b \in B : e \leq_1 b\}$*, such that* $\bigwedge A <_1 \bigwedge B$*, then (i) and (ii) hold:*
(i) $\Pi A \leq_1 \Pi B$
(ii) $\Sigma A \leq_1 \Sigma B$

Additionally, if $B \subseteq D$ *is an arbitrary set and* $A = B \cup E \cup \Pi E$ *for a nonempty set* $E = \{e \mid \forall b \in B : e \leq_1 b\}$*, such that* $\Pi A <_2 \Pi B$*, then (iii) and (iv) hold:*
(iii) $\bigwedge A \leq_2 \bigwedge B$
(iv) $\bigvee A \leq_2 \bigvee B$

Proof: The proof is analogous to the proof of 5.1.6. We only demonstrate the essential considerations for (i). Assume that $C = \bigwedge B \cup E \cup \bigwedge E$. Then $\Pi C = \bigwedge C \cdot \bigvee C = \bigwedge B \cdot \bigwedge C$ and $\bigwedge A \leq_1 \Pi C$ (by assumption). Consider $x := \Pi C \cdot \Pi B$. Obviously, $x \leq_2 C$ and $x \leq_2 B$. Therefore: $x \leq_2 A$. For similar reasons as in 5.1.6 we can justify: x is the greatest lower bound of A, and that is why we can establish the relation: $x = \Pi A$. Now we can formulate the inequality $\Pi A = \Pi C \cdot \Pi B \leq_1 \Pi C \vee \Pi B = \Pi B$ and the assertion $\Pi A \leq_1 \Pi B$ follows.
The assertions (ii)-(iv) are proven similarly. q.e.d.

The next Lemma is helpful to reduce the proof of Lemma 5.1.9 to the proofs of Lemmas 5.1.6 and 5.1.7. The following lemma is an immediate consequence of ordinary lattice theory and is *not* a result of the theory bilattices. Although the next claims are quite simple, they are important for the further argumentation.

Lemma 5.1.8 *Every bilattice* $\langle D, \leq_1, \leq_2 \rangle$ *satisfies the following conditions for arbitrary nonempty sets* A *and* B *where* $A, B \subseteq D$*:*

(i) $A \leq_1 B \rightarrow \bigwedge A \leq_1 \bigwedge B$ *and* $A \leq_1 B \rightarrow \bigvee A \leq_1 \bigvee B$
(ii) $A \leq_2 B \rightarrow \Pi A \leq_2 \Pi B$ *and* $A \leq_2 B \rightarrow \Sigma A \leq_2 \Sigma B$

Proof: (i) The proof requires only an argument concerning ordinary lattice theory. If $A \leq_1 B$, then by definition the following relations hold: $(\forall a \in A)(\exists b \in B) : a \leq_1 b$ and additionally $(\forall b \in B)(\exists a \in A) : a \leq_1 b$. Because of the second condition we can justify the claim: $\bigwedge A$ is a lower bound for every $b \in B$. Therefore we are able to conclude $\bigwedge A \leq_1 \bigwedge B$. Because of the first condition, we have immediately: $\bigvee B$ is an upper bound for every $a \in A$. Conclude: $\bigvee A \leq_1 \bigvee B$.

(ii) We can reason in a similar way as in (i) to get the desired result. q.e.d.

Now, it is possible to formulate a Lemma which enables us to give sufficient conditions for a bilattice to be interlaced. That result is essentially a generalization of Lemmas 5.1.6 and 5.1.7. Notice that Lemma 5.1.8 is crucially used here.

Lemma 5.1.9 *Let* $\mathbf{D} = \langle D, \leq_1, \leq_2 \rangle$ *be a bilattice, such that* $\mathbf{D}$ *satisfies the following properties:*

For all nonempty $A \subseteq D : \bigwedge A = \Pi A \wedge \Sigma A$ *and*
for all nonempty $A \subseteq D : \Pi A = \bigwedge A \cdot \bigvee A$ *and*
for all nonempty $A \subseteq D : \bigvee A = \Pi A \vee \Sigma A$ *and*
for all nonempty $A \subseteq D : \Sigma A = \bigwedge A + \bigvee A$

Then the bilattice $\langle D, \leq_1, \leq_2 \rangle$ *is interlaced.*

Proof: Assume the antecedent of the claim. Let A and B be arbitrary nonempty subsets of D, such that $A \leq_1 B$ holds. Because of Lemma 5.1.8 we simply have to check the following four claims (a)-(d):

(a) $[\bigwedge A = \bigwedge B$ and $\bigvee A = \bigvee B] \longrightarrow [\Pi A \leq_1 \Pi B$ and $\Sigma A \leq_1 \Sigma B]$
(b) $[\bigwedge A = \bigwedge B$ and $\bigvee A <_1 \bigvee B] \longrightarrow [\Pi A \leq_1 \Pi B$ and $\Sigma A \leq_1 \Sigma B]$
(c) $[\bigwedge A <_1 \bigwedge B$ and $\bigvee A = \bigvee B] \longrightarrow [\Pi A \leq_1 \Pi B$ and $\Sigma A \leq_1 \Sigma B]$
(d) $[\bigwedge A <_1 \bigwedge B$ and $\bigvee A <_1 \bigvee B] \longrightarrow [\Pi A \leq_1 \Pi B$ and $\Sigma A \leq_1 \Sigma B]$

Ad (a): Trivial, because of the assumption $\Pi A = \bigwedge A \cdot \bigvee A$ for all nonempty subsets $A \subseteq D$.
Ad (b): For the sets A and B it can be established: $\forall a \in A : \bigwedge B \leq_1 a$ and $a \leq_1 \bigvee B$. Consider now the set $C := A \cup \bigvee B \cup E$, with $E = \{e \mid \bigvee A \leq_1 e \leq_1 \bigvee B\}$. Obviously it holds: $\bigwedge C = \bigwedge B$ and $\bigvee C = \bigvee B$. With the premises $\Pi A = \bigwedge A \cdot \bigvee A$ and $\Sigma A = \bigwedge A + \bigvee A$ we have immediately $\Pi C = \bigwedge C \cdot \bigvee C = \bigwedge B \cdot \bigvee B = \Pi B$ and analogously the relation $\Sigma C = \bigwedge C + \bigvee C = \bigwedge B + \bigvee B = \Sigma B$. Because of the fact that C satisfies the conditions of Lemma 5.1.6 we are able to conclude: $\Pi A \leq_1 \Pi B$ and $\Sigma A \leq_1 \Sigma B$.
Ad (c): Similarly as in (b) it holds obviously: $\forall b \in B : \bigwedge A \leq_1 b$ and $b \leq_1 \bigvee A$. Now we consider the set C with $C := B \cup \bigwedge A \cup E$ where E is defined as follows: $E = \{e \mid \bigwedge A \leq_1 e \leq_1 \bigwedge B\}$. We can justify the relations $\bigwedge C = \bigwedge A$ and $\bigvee C = \bigvee A$. Moreover, it holds: $C \leq_1 B$. Therefore, we can verify $\Pi C = \bigwedge C \cdot \bigvee C = \bigwedge A \cdot \bigvee A$ and $\Sigma C = \bigwedge C + \bigvee C = \bigwedge A + \bigvee A = \Sigma A$. The set C satisfies the conditions of Lemma 5.1.7, so we have: $\Pi C \leq_1 \Pi B$ and $\Sigma C \leq_1 \Sigma B$ and therefore we are able to conclude the desired result (using the relations $\Pi C = \Pi A$ and $\Sigma C = \Sigma A$): $\Pi A \leq_1 \Pi B$ and $\Sigma A \leq_1 \Sigma B$.
Ad (d): The assertion is a consequence of (b) and (c): Consider a set C with $\bigwedge A = \bigwedge C$ and $\bigvee A <_1 \bigvee C = \bigvee B$. Then by (b) we have $\Pi A \leq_1 \Pi C$ and $\Sigma A \leq_1 \Sigma C$. Because of (c) we can conclude (using $\bigwedge C <_1 \bigwedge B$ and $\bigvee C = \bigvee B$): $\Pi C \leq_1 \Pi B$ and $\Sigma C \leq_1 \Sigma B$. Together we have $\Pi A \leq_1 \Pi B$ and $\Sigma A \leq_1 \Sigma B$ for all sets A and B with the property $\bigwedge A <_1 \bigwedge B$ and $\bigvee A <_1 \bigvee B$. Considering the cases (a)-(d), we get the following result:

$$\begin{array}{ll} \forall A, B \subseteq D : A \leq_1 B \longrightarrow \Pi A \leq_1 \Pi B & \text{and} \\ \forall A, B \subseteq D : A \leq_1 B \longrightarrow \Sigma A \leq_1 \Sigma B & \end{array}$$

The remaining cases, namely that it holds $A \leq_2 B \rightarrow \bigwedge A \leq_2 \bigwedge B$ and furthermore $A \leq_2 B \rightarrow \bigvee A \leq_2 \bigvee B$, are dual statements and can therefore be trivially established. q.e.d.

Using the above lemmas, it is possible to formulate Theorem 5.1.10 which states necessary and sufficient conditions for an arbitrary bilattice to be interlaced. In other words, the next theorem characterizes interlaced bilattices.

Theorem 5.1.10 *A bilattice $\langle D, \leq_1, \leq_2 \rangle$ is interlaced iff the following conditions hold:*

For all nonempty $A \subseteq D : \bigwedge A = \Pi A \wedge \Sigma A$ and
for all nonempty $A \subseteq D : \Pi A = \bigwedge A \cdot \bigvee A$ and
for all nonempty $A \subseteq D : \bigvee A = \Pi A \vee \Sigma A$ and
for all nonempty $A \subseteq D : \Sigma A = \bigwedge A + \bigvee A$

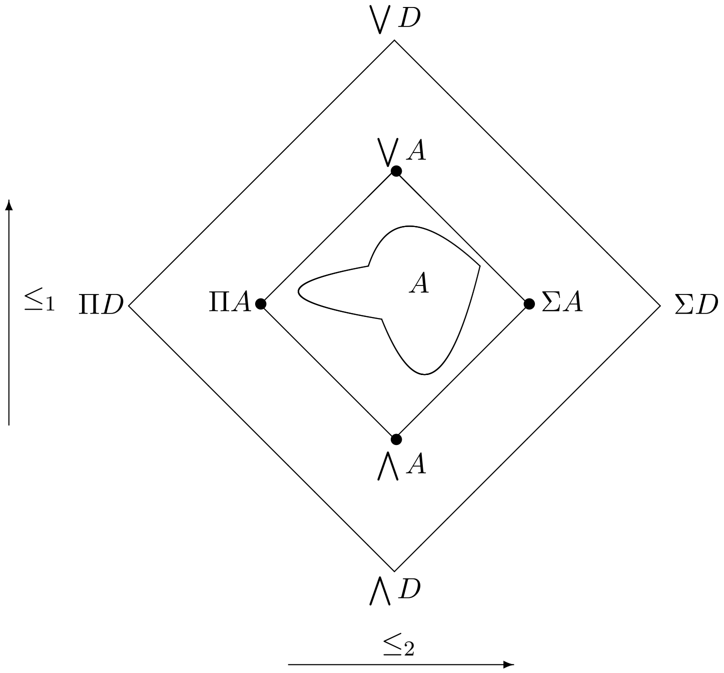
Proof: One direction of the claim is justified immediately by Lemma 5.1.5, the other direction follows directly from Lemma 5.1.9 q.e.d.

It is helpful to add some remarks, in order to provide a more intuitive picture of the characterization. Although Theorem 5.1.10 is quite abstract, it provides an intuitive picture of the properties of interlaced bilattices.

Remark 5.1.2 (i) Theorem 5.1.10 extends the claim of Fact 5.1.3(ii). Whereas Fact 5.1.3(ii) makes a claim about D the above theorem allows to generalize this claim to arbitrary subsets $A \subseteq D$ of a bilattice. An interesting feature of Fact 5.1.3(ii) is that Fitting proved his result in [Fi89] using transfinite inductions. The more general result 5.1.10 can be proven more elementary without transfinite techniques and the usage of ordinals.

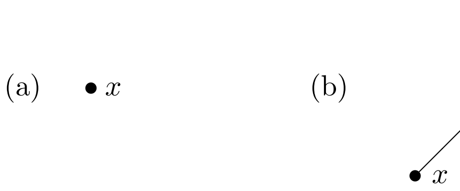
(ii) The interlacing condition connects the two order relations of a given bilattice. The characterization makes clear that this connection can be extended to an effective algorithm which calculates an infimum of a subset $A \subseteq D$ from the infimum and supremum of the alternative order relation. In other words: the order relation $\leq_i$ is fixed, if one knows precisely the behavior of the order relation $\leq_j$ for every subset $A \subseteq D$ ($i, j \in \{1, 2\}$).

(iii) The characterization seems to be quite abstract. In fact, the idea behind Theorem 5.1.10 is relatively simple. The following diagram helps to clarify the situation.

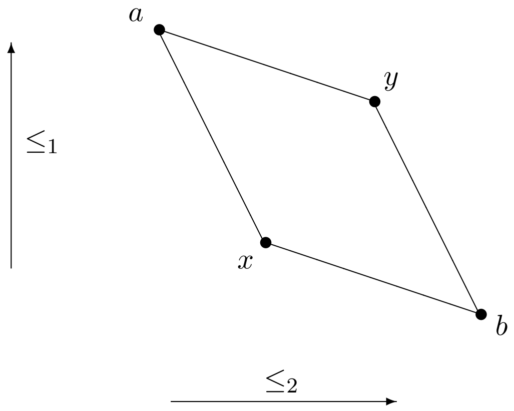


The diagram shows the behavior of an arbitrary subset A of a given bilattice $\mathbf{D} = \langle D, \leq_1, \leq_2 \rangle$. The supremum and infimum of A w.r.t. one of the two order relations can be calculated using the infimum and supremum of A w.r.t. the other order relation. In other words: given A, then we need only one order relation to calculate the supremum and infimum of the other extremal points of the other order relation. This is essentially the content of the interlacing condition.

(iv) Speaking metaphorically, Theorem 5.1.10 points out a property of interlaced bilattices that can be described with the (very intuitive) words 'every subset of an interlaced bilattice that includes the infima and suprema of both order relations is 'rhombus-like' and every 'rhombus-like' bilattice is interlaced'. Here, the term 'rhombus-like' refers to the characterization properties of Theorem 5.1.10. If we consider a finite interlaced bilattice (we require furthermore that this bilattice is representable in a 'double-Hasse-diagram'), we can associate this bilattice with a bilattice that is composed of several rhombuses. Each such rhombus need not to be the trivial rhombus with four elements (`FOUR`); but each rhombus can be represented by a generalized form of rhombuses. The limits of such a metaphorical illustration are obvious. For example, the following 'rhombuses' (a) and (b) are also interlaced bilattices: both do not have the (intuitive) diagrammatic rhombus-like property, but both satisfy the required conditions for an interlaced bilattice (provided that in example (b) we interpret $\leq_1 = \leq_2$ and we assume furthermore that $x \leq_i y$ for $i \in \{1, 2\}$).



(v) It would be a nice property of interlaced bilattices, if each pair of elements of a given interlaced bilattice was comparable at least with respect to one order relation (later we will call this property 'bilinearity'). Unfortunately, this is not necessarily the case, as can be seen in the following diagram. Although the bilattice in the diagram is interlaced, x and y are neither comparable with respect to $\leq_1$ nor to $\leq_2$.



In the next section, we will prove a second characterization theorem that states sufficient and necessary conditions for a particular subclass of interlaced bilattices.

5.2 A Second Characterization Theorem

In order to give characterizations of the interlacing property of subclasses of bilattices, we restrict our consideration to bilinear bilattices and distributive bilattices in the following two sections. First, we introduce the important concept of bilinearity in the next definition. This concept was originally introduced by Melvin Fitting in [Fi94].

Definition 5.2.1 *A bilattices* $\mathbf{D} = \langle D, \leq_1, \leq_2 \rangle$ *is called bilinear if each two elements* $x \in D$ *and* $y \in D$ *are comparable with each other at least with respect to one of the two order relations* $\leq_1$ *and* $\leq_2$.

The idea of bilinearity is that each pair of elements of D are connected via an order relation. At this point, it is necessary to define the notion of a bifilter, originally introduced in [ArAv96] which provides a new characterization of a certain subclass of interlaced bilattices. We begin our argumentation with some definitions that are direct generalizations of the corresponding concepts in ordinary lattice theory.

Definition 5.2.2 *(i) Let* $\langle D, \leq_1, \leq_2 \rangle$ *be a bilattice. A set* $F \subseteq D$ *is called a bifilter, if (a) and (b) hold:*
(a) $\forall a, b \in D : a \wedge b \in F \;\Leftrightarrow\; a \in F$ *and* $b \in F$
(b) $\forall a, b \in D : a \cdot b \in F \;\Leftrightarrow\; a \in F$ *and* $b \in F$

(ii) A bifilter F *of a bilattice* $\langle D, \leq_1, \leq_2 \rangle$ *is called prime, if the conditions (c) and (d) hold:*
(c) $\forall a, b \in F : a \vee b \in F \;\Leftrightarrow\; a \in F$ *or* $b \in F$
(d) $\forall a, b \in F : a + b \in F \;\Leftrightarrow\; a \in F$ *or* $b \in F$

Bifilters are straightforward generalizations of ordinary filters in lattice theory (or other algebraic theories). Similarly, the same is true for the concept of prime bifilters. An additional remark is helpful at this point: An ultrabifilter, according to [ArAv96], needs additional operations defined on a bilattice. An ultrabifilter is defined by additional operations which are interpretable as certain kinds of negations (from a logical point of view) or certain isomorphisms (from a more algebraic point of view). A similar situation exists in lattice theory. The maximality condition of an ultrafilter uses essentially a negation. Because of the fact that in the theory of bilattices there are two possible negations, we need to consider two lattice isomorphisms. In the remaining sections of this chapter, we do not work with ultrabifilters and refer the interested reader to the paper [ArAv96] for more information concerning this concept.

The next Lemma describes a property of interlaced bilattices. It is not dependent on the bilinearity condition of bilattices.

Lemma 5.2.3 *Let* $\langle D, \leq_1, \leq_2 \rangle$ *be an interlaced bilattice. Then it holds for all* $x \in D$:

$$x \leq_1 \Pi D \;\Leftrightarrow\; x \leq_2 \bigwedge D$$

Proof: The proof is an easy application of the interlacing condition of $\langle D, \leq_1, \leq_2 \rangle$. For one direction we assume that $x \leq_1 \Pi D$. Then, the relations $\Pi D \leq_2 \bigwedge D$ and $x \leq_2 x$ hold obviously. Using the interlacing condition of $\langle D, \leq_1, \leq_2 \rangle$ we can deduce: $x = \Pi D \wedge x \leq_2 \bigwedge D \wedge x = \bigwedge D$. For the other direction we assume $x \leq_2 \bigwedge D$. Clearly, it holds: $\bigwedge D \leq_1 \Pi D$ and $x \leq_1 x$. Again we use the interlacing condition and get: $x = \bigwedge D \cdot x \leq_1 \Pi D \cdot x = \Pi D$.

Together: $x \leq_2 \Pi D \Leftrightarrow x \leq_1 \bigwedge D$. q.e.d.

Remark 5.2.1 It is clear that there is a dual relation of the claim of Lemma 5.2.3 with respect to the points ΣD and $\bigvee D$. This is a direct consequence of the symmetric structure of bilattices. Because of the fact that this dual relation is not needed in the further considerations, we skip it.

The next Lemma states a sufficient condition for a set X to be a chain in D for a given billinear bilattice.

Lemma 5.2.4 *Let $\langle D, \leq_1, \leq_2 \rangle$ be a bilinear interlaced bilattice and let*

$$X := \{x \mid \bigwedge D \leq_1 x \leq_1 \Pi D\}$$

be a subset of D. Then X is a chain with respect to $\leq_1$ and $\leq_2$.

Proof: We have to prove that every two elements of X are comparable with respect to $\leq_1$ and $\leq_2$. Let x and y be two arbitrary elements of X. We show that the following equivalence holds: $x \leq_1 y \leq_2 x$. Assume $x \leq_1 y$. Obviously, we have $x \leq_1 x$ and $y \leq_1 \Pi D$. Using Lemma 5.2.3, we can verify the two relations $x \leq_2 x$ and $y \leq_2 \bigwedge D$ and with the interlacing condition of $\langle D, \leq_1, \leq_2 \rangle$ we get: $y = x \vee y \leq_2 x \vee \bigwedge D = x$. For the other direction the reasoning is similar. Together we have: $x \leq_1 y \Leftrightarrow y \leq_2 x$. Because $\langle D, \leq_1, \leq_2 \rangle$ is bilinear, we can conclude: X is a chain with respect to both order relations. q.e.d.

Now, we are able to prove a Lemma which provides necessary conditions for a (non-trivial) bilinear bilattice that is interlaced.

Lemma 5.2.5 *Let $\langle D, \leq_1, \leq_2 \rangle$ be a bilinear non-trivial interlaced bilattice.*[1] *Then the subset $E \subseteq D$ that is defined as follows:*

$$E := D - \{x \mid \bigwedge D \leq_1 x \leq_1 \Pi D\}$$

is a prime bifilter of $\langle D, \leq_1, \leq_2 \rangle$.

Proof: We have to check that E satisfies the conditions of Definition 5.2.2(i). First, we show that the following two conditions hold:

$$\text{(a) } \forall a, b \in D : a \wedge b \in E \ \Leftrightarrow\ a \in E \text{ and } b \in E$$

$$\text{(b) } \forall a, b \in D : a \cdot b \in E \ \Leftrightarrow\ a \in E \text{ and } b \in E$$

Ad (a): Assume that $a \wedge b \in E$. We can justify either $a \wedge b \not\leq_1 \Pi D$ or $a \wedge b \not\leq_2 \bigwedge A$, according to the definition of E. Because of Lemma 5.2.3 we have the equivalence $\neg(a \wedge b \not\leq_2 \bigwedge D) \Leftrightarrow \neg(a \wedge b \not\leq_1 \Pi D)$ and therefore we get

[1] The term 'non-trivial' means that $\langle D, \leq_1, \leq_2 \rangle$ is not a chain: $\{x \mid \bigwedge D \leq_1 x \leq_1 \Pi D\} \neq D$.

the implication: $a \wedge b \in E \Rightarrow \neg(a \wedge b \leq_1 \Pi D)$. Because it holds $a \wedge b \leq_1 a$ and $a \wedge b \leq_1 b$ we have a fortiori $\neg(a \leq_1 \Pi D)$ and $\neg(b \leq_1 \Pi D)$. Thus: $a \in E$ and $b \in E$, according to the definition of E.
Now, assume again that $a \in E$ and $b \in E$. Because $\langle D, \leq_1, \leq_2\rangle$ is bilinear we claim: a and b are comparable with respect to one of the two order relations (at least). Without loss of generality we can distinguish two cases: $a \leq_1 b$ or $a \leq_2 b$. If it holds $a \leq_1 b$ we have immediately $a = a \wedge b \in E$. On the other hand, if $a \leq_2 b$, then we get $a \cdot b \in E$ and $a + b \in E$ as well. Assume that it holds $a \wedge b \notin E$. Then by definition: $a \cdot b \leq_2 a \wedge b \leq_2 \bigwedge D$ and together with Lemma 5.2.4 we have $a \cdot b \notin E$ which is a contradiction. Therefore we can conclude that $a \wedge b \in E$ holds.

Ad (b): The assertion $\forall a, b \in D : a \cdot b \in E \Leftrightarrow a \in E$ and $b \in E$ is the dual of the above reasoning.

It remains to show that (c) and (d) holds:

$$\text{(c) } \forall a, b \in E : a \vee b \in E \;\Leftrightarrow\; a \in E \text{ or } b \in E$$

$$\text{(d) } \forall a, b \in E : a + b \in E \;\Leftrightarrow\; a \in E \text{ or } b \in E$$

For one direction of the first claim, we have to show that the following implication holds: $a \vee b \in E \Rightarrow (a \in E$ or $b \in E)$. We show the contraposition: If $a \notin E$ and $b \notin E$, then by definition of E we get $a \in D - E = \{x \mid \bigwedge D \leq_1 x \leq_1 \Pi D\}$ and $b \in D - E$. But $D - E$ is a chain with respect to both order relations (because of Lemma 5.2.4), therefore we have immediately the fact: $a \vee b \in D - E$ i.e. $a \vee b \notin E$.
The other direction is obvious: If $a \in E$ or $b \in E$, then $a \vee b \in E$, because it holds: if $a \vee b \notin E$, then by definition $a \vee b \leq_1 \Pi D$. Then we have a fortiori $a \leq_1 \Pi D$ and $b \leq_1 \Pi d$ which means: $a \notin E$ and $b \notin E$. Together we conclude: $a \vee b \in E \Leftrightarrow a \in E$ or $b \in E$.

The relation $a + b \in E \Leftrightarrow a \in E$ or $b \in E$ is the dual statement of the above reasoning. Hence, the claim follows immediately from the above considerations.
q.e.d.

Now, we can prove the second characterization theorem. It gives us a possibility to characterize the class of nontrivial finite bilinear interlaced bilattices.

Theorem 5.2.6 *Let $\langle D, \leq_1, \leq_2\rangle$ be a non-trivial bilinear finite bilattice. Then it holds: $\langle D, \leq_1, \leq_2\rangle$ is interlaced iff $E \subset D$ with*

$$E := D - \{x \mid \bigwedge D \leq_1 x \leq_1 \Pi D\}$$

is a prime bifilter of $\langle D, \leq_1, \leq_2\rangle$, $D - E$ satisfies the interlacing conditions of $\langle D, \leq_1, \leq_2\rangle$ and the set $\{x \mid \bigwedge D \leq_1 x \leq_1 \Pi D\}$ is a chain with respect to both order relations.

Proof: "$\Rightarrow$" If $\langle D, \leq_1, \leq_2 \rangle$ is a finite non-trivial bilinear interlaced bilattice, then E is a prime bifilter of $\langle D, \leq_1, \leq_2 \rangle$ because of Lemma 5.2.5. Notice that $\langle E, \leq_1, \leq_2 \rangle$ itself is an interlaced bilattice: This holds, because all infima with respect to both order relations are elements of E: Let $A = \{a_1, a_2, ..., a_n\} \subseteq E$ be an arbitrary set. Then consider the following computation of $\bigwedge A$:

$$\begin{array}{llll} a_0 \wedge a_1 = c_1 \\ c_1 \wedge a_2 = c_2 \\ \vdots & \vdots & \vdots & \vdots \\ c_{n-1} \wedge a_n = c_n \end{array}$$

Obviously, all c_is $(i \leq n)$ are elements of E, because E is a prime bifilter of $\langle D, \leq_1, \leq_2 \rangle$. For the infimum operation with respect to $\leq_2$ we reason in a similar way. Moreover, the set

$$\{x \mid \bigwedge D \leq_1 x \leq_1 \Pi D\}$$

is a chain with respect to both order relations (because of Lemma 5.2.4).

"$\Leftarrow$" Trivial. q.e.d.

We can state as an easy consequence the following corollary. This corollary is a representation statement for the class of finite bilinear interlaced bilattices.

Corollary 5.2.7 *Assume $\langle D, \leq_1, \leq_2 \rangle$ is a bilattice. Then the following equivalence holds: $\langle D, \leq_1, \leq_2 \rangle$ is a finite bilinear interlaced bilattice iff $\langle D, \leq_1, \leq_2 \rangle$ is isomorphic to the product of two chains $\langle \{0, 1, 2, ..., n\}, \leq \rangle$ and $\langle \{0, 1, 2, ..., m\}, \leq \rangle$ where $\leq$ is the standard order on the natural numbers.*

Proof: "$\Leftarrow$" Assume that $\langle D, \leq_1, \leq_2 \rangle$ is a finite, bilinear and interlaced bilattice. We have to show that there are two chains $A = \langle \{0, 1, 2, ..., n\}, \leq \rangle$ and $B = \langle \{0, 1, 2, ..., m\}, \leq \rangle$, such that the product $A \times B$ is isomorphic to $\langle D, \leq_1, \leq_2 \rangle$. Using Lemma 5.2.4 we know that $\{x \mid \bigwedge D \leq_1 x \leq_1 \Pi D\}$ is a chain with respect to both order relations and $D^{-1} = D - \{x \mid \bigwedge D \leq_1 x \leq_1 \Pi D\}$ is again a finite bilinear interlaced bilattice. Notice that a finite application of this reduction (for a certain natural number n) yields a bilattice $D^{-n} = D^{-n+1} - \{x \mid \bigwedge D^{-n+1} \leq_1 x \leq_1 \Pi D^{-n+1}\}$ that is a chain with respect to both order relations. Clearly, D^{-n} is isomorphic to a product of two chains, namely $D^{-n} \cong D^{-n} \times \langle \{0\}, \leq \rangle \cong \langle \{0, 1, 2, ..., m\}, \leq \rangle \times \langle \{0\}, \leq \rangle$. Now, it is clear that $D^{-n+1} \cong \langle \{0, 1, 2, ..., m\}, \leq \rangle \times \langle \{0, 1\}, \leq \rangle$ and so on. Finally, we get $D \cong \langle \{0, 1, 2, ..., m\}, \leq \rangle \times \langle \{0, 1, 2, ..., m\}, \leq \rangle$ which is the desired result.

"$\Rightarrow$" Assume that $\langle \{0, 1, 2, ..., n\}, \leq \rangle$ and $\langle \{0, 1, 2, ..., m\}, \leq \rangle$ are two chains. Let $\langle \{0, 1, 2, ..., n\} \times \{0, 1, 2, ..., m\}, \leq_1, \leq_2 \rangle$ be the bilattice resulting from the product of the two chains. Because of Fact 5.1.3(iii), this bilattice is interlaced. Finiteness is clear, and bilinarity is a direct consequence of the definition of the induced order relations. q.e.d.

Characterization Theorem 5.2.6 clarfies that the bilinearity condition of bilattices is a relatively strong restriction of bilattices. Essentially, bilinear interlaced bilattices are products of chains (provided the bilatice is finite). That is the upshot of Corollary 5.2.7, the representation theorem for this class of interlaced bilattices.

In the next section, we will prove a third characterization of a certain subclass of interlaced bilattices. We will correlate interlaced bilattices with distributive bilattices and we will see that the connection between the two classes is the bilinearity condition.

5.3 A Third Characterization Theorem

It is well known by the work of Melvin Fitting (compare [Fi89]) that every distributive bilattice is also interlaced. The converse is not true in general. According to the following Theorem 5.3.1, we can specify an additional property (i.e. we specify a subclass of bilattices) in order to get an equivalence between interlaced bilattices and distributive bilattices. If we consider only bilinear bilattices, then the class of interlaced bilattices and the class of distributive bilattices are coextensional. In other words: every bilinear bilattice which is interlaced is also distributive and vice versa.

Theorem 5.3.1 *Let $\langle D, \leq_1, \leq_2 \rangle$ be a bilinear bilattice. Then it holds:* *$\langle D, \leq_1, \leq_2 \rangle$ is interlaced iff $\langle D, \leq_1, \leq_2 \rangle$ is distributive.*

Proof: "$\Rightarrow$" We have to prove that a bilinear interlaced bilattice satisfies the following 12 equations for all $a, b, c \in D$:

(i) $a \wedge (b \cdot c) = (a \wedge b) \cdot (a \wedge b)$
(ii) $a \wedge (b + c) = (a \wedge b) + (a \wedge b)$
(iii) $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge b)$
(iv) $a \vee (b \cdot c) = (a \vee b) \cdot (a \vee c)$
(v) $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$
(vi) $a \vee (b + c) = (a \vee b) + (a \vee c)$
(vii) $a \cdot (b \wedge c) = (a \cdot b) \wedge (a \cdot c)$
(viii) $a \cdot (b \vee c) = (a \cdot b) \vee (a \cdot c)$
(ix) $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$
(x) $a + (b \cdot c) = (a + b) \cdot (a + c)$
(xi) $a + (b \vee c) = (a + b) \vee (a + c)$
(xii) $a + (b \wedge c) = (a + b) \wedge (a + c)$

A remark concerning the following considerations is helpful: each of the justifications of (i)-(xii) is very similar in its structure. In particular, the proofs for the claims (i)-(xii) are essentially a matter of case-checking. Although the reasoning is very straightforward (in general), the reader could find it quite tedious. We do not prove the theorem explicitly and in all details, but rather sketch the proof. The reader is referred to [Ku99] for a more complete and explicit proof.

Ad(i): We have to prove that the equation $a \wedge (b \cdot c) = (a \wedge b) \cdot (a \wedge b)$ holds. One direction is obvious: Because of the valid inequalities $a \wedge b \leq_1 b$ and $a \wedge c \leq_1 c$, we have immediately $(a \wedge b) \cdot (a \wedge c) \leq_1 (b \cdot c)$ (using the interlacing condition). Furthermore, because of $a \wedge b \leq_1 a$ and the fact that $a \wedge c \leq_1 a$ we get the valid formula $(a \wedge b) \cdot (a \wedge c) \leq_1 a$ and together: $(a \wedge b) \cdot (a \wedge c) \leq_1 a \wedge (b \cdot c)$.[2]
The other direction requires a distinction in nine subcases (a) - (i). We check all possible relations between $b \cdot c$, and the two elements b and c in the following.

(a) Assume $b \cdot c \leq_1 b$ and $b \cdot c \leq_1 c$. Using the relation $a \leq_1 a$ we have: $a \wedge (b \cdot c) \leq_1 a \wedge b$ and $a \wedge (b \cdot c) \leq_1 a \wedge c$. With the interlacing condition we get the desired result: $a \wedge (b \cdot c) \leq_1 (a \wedge b) \cdot (a \wedge c)$.
(b) Assume $b \leq_1 b \cdot c$ and $b \cdot c \leq_1 c$. Then, we can justify the two relations $a \wedge (b \cdot c) \leq_1 a \wedge c$ and $a \wedge b \leq_1 a \wedge b$. This gives us the relation $(a \wedge (b \cdot c)) \cdot (a \wedge b) \leq_1 (a \wedge b) \cdot (a \wedge c)$. Because $b \cdot c \leq_2 b$ (and therefore $a \wedge (b \cdot c) \leq_2 a \wedge b$), we can justify $(a \wedge (b \cdot c)) \cdot (a \wedge b) = a \wedge (b \cdot c)$. But then: $a \wedge (b \cdot c) \leq_1 (a \wedge b) \cdot (a \wedge c)$.
(c) Assume $b \cdot c \leq_1 b$ and $c \leq_1 b \cdot c$. Similar reasoning as in (b) provides the desired result: $a \wedge (b \cdot c) \leq_1 (a \wedge b) \cdot (a \wedge c)$.
(d) Assume $b \leq_1 b \cdot c$ and $c \leq_1 b \cdot c$. Because $b \cdot c$ is an upper bound of b and c with respect to $\leq_1$, we have $b \vee c \leq_1 b \cdot c$. This implies (using Lemma 5.1.4) that $b \vee c = b \cdot c$. Because of the validity of the following two relations $b + c \leq_1 b \cdot c$ and $b + c \leq_1 b + c$, we get $b + c \leq_1 b \wedge c$ which means that $b + c = b \wedge c$. Obviously, $\{b, c\}$ builds a chain with respect to both order relations. Without loss of generality assume $c \leq_2 b$. Then, we can justify $a \wedge (b \cdot c) \leq_1 a \wedge c$ and $a \wedge b \leq_1 a \wedge b$. With the interlacing condition we can deduce the relation $(a \wedge (b \cdot c)) \cdot (a \wedge b) \leq_1 (a \wedge b) \cdot (a \wedge c)$ and this is equivalent to the desired inequality $a \wedge (b \cdot c) \leq_1 (a \wedge b) \cdot (a \wedge c)$.
(e) Let b and $b \cdot c$ be not comparable with respect to $\leq_1$ and assume further that $b \cdot c \leq_1 c$. Because of the valid two relations $a \wedge (b \cdot c) \leq_1 a \wedge c$ and $a \cdot b \leq_1 a \wedge b$ we get immediately $(a \wedge (b \cdot c)) \cdot (a \wedge b) \leq_1 (a \wedge b) \cdot (a \wedge c)$. This relation is equivalent to $a \wedge (b \cdot c) \leq_1 (a \wedge b) \cdot (a \wedge c)$ which is the desired result.
(f) Assume $b \cdot c \leq_1 b$ and let c and $b \cdot c$ be not comparable with respect to $\leq_1$. The same reasoning as in (e) guarantees that it holds: $a \wedge (b \cdot c) \leq_1 (a \wedge b) \cdot (a \wedge c)$.
(g) Let b and $b \cdot c$ be not comparable with respect to $\leq_1$ and assume further that $c \leq_1 b \cdot c$. Assume b and c are comparable with respect to $\leq_1$, then $b = b \vee c$ or $b = b \wedge c$, and therefore b is comparable with the element $b \cdot c$ with respect to $\leq_1$. But that contradicts our assumption. So, b and c is only comparable with respect to $\leq_2$ (notice that at this point the bilinearity condition of $\langle D, \leq_1, \leq_2 \rangle$ is essentially used). Without loss of generality we assume $b \cdot c = c$ and $b + c = b$. Then, the relations $a \wedge (b \cdot c) \leq_1 (a \wedge c)$ and $a \wedge b \leq_1 a \wedge b$ hold. Using the interlacing condition we have immediately the relation $(a \wedge (b \cdot c)) \cdot (a \wedge b) \leq_1 (a \wedge b) \cdot (a \wedge c)$.

[2] Notice that this direction corresponds to the 'easy' direction in ordinary lattice theory.

This inequality is equivalent to the desired relation $a \wedge (b \cdot c) \leq_1 (a \wedge b) \cdot (a \wedge c)$.
(h) Assume $b \leq_1 b \cdot c$ and let c and $b \cdot c$ be not comparable with respect to $\leq_1$. Then, we reason similarly as in case (g).
(i) Assume both, namely that b and $b \cdot c$ and c and $b \cdot c$ are not comparable with respect to $\leq_1$. If it were the case that b and c are comparable with respect to $\leq_1$, then $b \leq_1 b \cdot c$ or $b \cdot c \leq_1 b$ (justified by Lemma 5.1.4) and we would have a contradiction with our assumption. If it were the case that b and c are comparable with respect to $\leq_2$, then we would have $b = b \cdot c$ or $c = b \cdot c$, and again we would have a contradiction. So, this case is not possible in an interlaced bilattice which is bilinear.
Conclusion: it holds $a \wedge (b \cdot c) \leq_1 (a \wedge b) \cdot (a \wedge c)$ and together with the inequality of the other direction above, we have justified the desired result: $a \wedge (b \cdot c) = (a \wedge b) \cdot (a \wedge c)$.

Ad (ii) First, we show the straightforward direction $(a \wedge b) + (a \wedge c) \leq_1 a \wedge (b + c)$. Obviously it holds $a \wedge b \leq_1 a$ and $a \wedge c \leq_1 a$, and using the interlacing property we get: $(a \wedge b) + (a \wedge c) \leq_1 a$. Because of the fact that it holds $a \wedge b \leq_1 b$ and $a \wedge c \leq_1 c$, we get immediately $(a \wedge b) + (a \wedge c) \leq_1 b + c$. Together we have: $(a \wedge b) + (a \wedge c) \leq_1 a \wedge (b + c)$.

The proof for the other direction $a \wedge (b + c) \leq_1 (a \wedge b) + (a \wedge c)$ is more complex. Again we have to consider nine subcases (a) - (i):

(a) Assume $b + c \leq_1 b$ and $b + c \leq_1 c$. Together with the trivial inequality $a \leq_1 a$ we get immediately on the one hand the inequality $a \wedge (b + c) \leq_1 a \wedge b$ and on the other hand the inequality $a \wedge (b + c) \leq_1 a \wedge c$. Using the interlacing property we get: $a \wedge (b + c) \leq_1 (a \wedge b) + (a \wedge c)$.
(b) Assume $b+c \leq_1 b$ and $c \leq_1 b+c$. On the one hand we have $a \wedge (b+c) \leq_1 a \wedge b$, and on the other hand we have trivially the inequality $a \wedge c \leq_1 a \wedge c$. That leads us to the inequality $(a \wedge (b + c)) + (a \wedge c) \leq_1 (a \wedge b) + (a \wedge c)$. Now, $c \leq_2 b + c$ and therefore we get $(a \wedge c) \leq_2 a \wedge (b + c)$. That is why we can conclude the desired relation: $a \wedge (b + c) \leq_1 (a \wedge b) + (a \wedge c)$.
(c) Assume $b \leq_1 b + c$ and $b + c \leq_1 c$. We have symmetrical conditions as in subcase (b), therefore a similar reasoning (with substituting b for c) guarantees the desired inequality $a \wedge (b + c) \leq_1 (a \wedge b) + (a \wedge c)$.
(d) Assume $b \leq_1 b + c$ and $c \leq_1 b + c$. With Lemma 5.1.4 we get $b + c = b \vee c$. Then we have obviously: $b \cdot c \leq_1 b + c$ and $b \cdot c \leq_1 b \cdot c$ and therefore it holds: $b \cdot c \leq_1 (b + c) \wedge (b \cdot c) = b \wedge c$ (by Lemma 5.1.5). We conclude: $b \cdot c = b \wedge c$. Obviously $\{b, c\}$ forms a chain with respect to both order relations. Without loss of generality assume $b \leq_2 c$. Then we have: $a \wedge (b+c) \leq_1 a \wedge c$ and $a \wedge b \leq_1 a \wedge b$. With the interlacing condition we get: $(a \wedge (b+c)) + (a \wedge b) \leq_1 (a \wedge b) + (a \wedge c)$. Because of the fact that $b \leq_2 b + c$ (and therefore $a \wedge b \leq_2 a \wedge (b + c)$) we conclude that it holds: $a \wedge (b + c) \leq_1 (a \wedge b) + (a \wedge c)$.
(e) Let b and $b+c$ be not comparable with respect to $\leq_1$ and assume $b+c \leq_1 c$. Because of the two relations $a \wedge (b + c) \leq_1 a \wedge c$ and $a \wedge b \leq_1 a \wedge b$, we get the relation $(a \wedge (b + c)) + (a \wedge b) \leq_1 (a \wedge b) + (a \wedge c)$. This is equivalent to the

relation: $a \wedge (b + c) \leq_1 (a \wedge b) + (a \wedge c)$.
(f) Assume $b + c \leq_1 b$ and let c and $b + c$ be not comparable with respect to $\leq_1$. Again, we reason in the similar way as in (e).
(g) Let b and $b+c$ be not comparable with respect to $\leq_1$ and assume $c \leq_1 b+c$. It is not possible that b and c are comparable with respect to $\leq_1$, because else we would get $b \leq_1 b + c$ or $b + c \leq_1 b$ which is inconsistent with our assumption. Therefore b and c must be comparable with respect to $\leq_2$ (at this step we use the bilinearity assumption). Because b and $b + c$ are incomparable with respect to $\leq_1$ we get $b = b \cdot c$ and $c = b + c$. Therefore: $a \wedge b \leq_1 a \wedge b$ and $a \wedge (b + c) \leq_1 a \wedge c$. Using the interlacing condition we get the relation: $(a \wedge (b + c)) + (a \wedge b) \leq_1 (a \wedge b) + (a \wedge c)$. As above this is equivalent to: $a \wedge (b + c) \leq_1 (a \wedge b) + (a \wedge c)$.
(h) Assume $b \leq_1 b + c$ and let c and $b + c$ be not comparable with respect to $\leq_1$. Similar reasoning as in subcase (g) guarantees that it holds: $a \wedge (b + c) \leq_1 (a \wedge b) + (a \wedge c)$.
(i) Let b and $b + c$, and c and $b + c$ be not comparable with respect to $\leq_1$. Assume b and c are comparable with respect to $\leq_1$, then it would be the case that $b \leq_1 b + c$ or $c \leq_1 b + c$ which is impossible. Assume on the other hand that b and c are comparable with respect to $\leq_2$, then we have $b = b + c$ or $c = b + c$, which is impossible, either. Hence, this situation cannot occur.

We can state the following conclusion: $a \wedge (b + c) \leq_1 (a \wedge b) + (a \wedge c)$ and with the inequality of the other direction ('easy' direction) we have the equality: $a \wedge (b + c) = (a \wedge b) + (a \wedge c)$.

Ad (iii)-(xii): It is a tedious and boring work to check all the remaining cases. We refer the interested reader to [Ku99] for a complete proof of this direction.

"$\Leftarrow$" Compare Fact 5.1.3(i). q.e.d.

Remark 5.3.1 (i) The proof of Theorem 5.3.1 is not very elegant. Although the reasoning is relatively simple in character (one has only to check a certain number of cases and subcases), it is a tedious work to write down this proof. The author was not able to simplify the presented proof or to find an alternative (and easier) argument.

(ii) Theorem 5.3.1 illuminates the crucial property connecting interlaced bilattices and distributive bilattices: this crucial property is the bilinearity condition. Restricting the discourse to bilinear bilattices, distributivity and the interlacing property are equivalent. Differences between these concepts arise, if there are points that are incomparable with respect to both order relations. Notice that Theorem 5.3.1 does not include any restriction with respect to the cardinality of the bilattice. It is completely general and does not use higher order concepts like ordinal numbers.

(iii) Notice that Corollary 5.2.7 is in a certain sense a finite version of Theorem 5.3.1. The fact that finite, bilinear interlaced bilattices are products of (finite) chains implies that these bilattices are also distributive.[3] The special character of Corollary 5.2.7 is naturally generalized by the Theorem 5.3.1.

(iv) The bilinearity condition is a relatively strong restriction. Most examples we saw in this chapter and in Chapter 4 were not bilinear. The classical examples of bilinear bilattices $\mathbf{D} = \langle D, \leq_1, \leq_2 \rangle$ are products of chains (compare Definition 4.2.4) and bilattices that are build out of the simple bilattice `FOUR` (compare Example 4.2.3(ii) for the definition of `FOUR`).

One could imagine to consider another kind of distributivity condition. In general, it makes a difference whether one considers the finite form of distributivity (as for example the condition $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$) or the infinite form as can be represented as follows (where $B \subseteq D$ is a countable set):

$$\bigvee(\bigwedge\{a, b_i\}_{b_i \in B}) = a \wedge (\bigvee B)$$

It is clear that the infinite version implies the finite one. The contrary is not generally true in the theory of lattices (as well as in topological spaces where the intersection of infinitely many open sets is not necessarily in the topological space). The crucial condition in the theory of lattices is completeness. For the class of bilattices completeness is assumed. Therefore, the finite distributivity version implies even the infinite one. The following proposition states that fact.

Proposition 5.3.2 *Assume* $\mathbf{D} = \langle D, \leq_1, \leq_2 \rangle$ *is a countable bilattice. Then it holds: if* $\mathbf{D}$ *is finitely distributive then* $\mathbf{D}$ *is infinitely distributive as well.*

Proof: The crucial point in the proof is the completeness condition of the two lattices $\langle D, \leq_1 \rangle$ and $\langle D, \leq_2 \rangle$. We show only that it holds: $\bigvee(\bigwedge\{a, b_i\}_{b_i \in \mathbb{N}}) = a \wedge (\bigvee B)$. For one direction notice that it holds

$$\forall i \in \mathbb{N} : a \wedge b_i \ \leq_1 \ a \wedge (b_1 \vee b_2 \vee ...)$$

Hence, we can conclude that it holds:

$$(a \wedge b_1) \vee (a \wedge b_2) \vee ... \ \leq_1 \ a \wedge (b_1 \vee b_2 \vee ...)$$

For the other direction we have (using the assumed finite distributivity condition)

$$\forall n \in \mathbb{N} : a \wedge (b_1 \vee b_2 \vee ... \vee b_n) \leq_1 (a \wedge b_1) \vee (a \wedge b_2) \vee ... \vee (a \wedge b_n)$$

Because of completeness we can deduce that it holds:

$$\bigvee(\bigwedge\{a, b_i\}_{b_i \in B}) = a \wedge (\bigvee B)$$

[3] This is easy to check.

That suffices to show the proposition. q.e.d.

Remark 5.3.2 The above Proposition 5.3.2 shows that we can work with finite distributivity instead of infinite distributivity without loosing important properties (at least for the countable case). The crucial point is the completeness of the underlying lattices. Working in the class of pre-bilattices where completeness is not required the distinction between finite and infinite distributivity becomes important.

We finish this section of the third characterization of subclasses of interlaced bilattices with these remarks. The next section is devoted to some considerations concerning possible negations (lattice homomorphisms) in bilattices in general. Although this is not absolutely important for Kripke's fixed point approach it is helpful for a better understanding of the concept of bilattices.

5.4 Negation

In this section, we will consider the possibilities to introduce a negation on interlaced bilattices. We remind the reader that the original definition of a bilattice by Ginsberg includes a lattice homomorphism preserving the $\leq_I$ order and inverts the $\leq_T$ order relation (cf. Definition 4.2.3). We do not want to restrict our considerations to truth orders and information orders. We use simply the relations $\leq_1$ (that corresponds closely to the information order) and $\leq_2$ (that corresponds to the truth order in Definition 4.2.3). Therefore, we assume that the negation $\neg$ preserves the $\leq_1$ order relation and inverts the $\leq_2$ order relation.

We use the following notation for an arbitrary subset $A \subseteq D : \neg A = \{\neg a_i\}_{a_i \in A}$. This notation will simplify the following consideration. First, notice that the following fact holds:

Fact 5.4.1 *For interlaced bilattices it is not in general possible to introduce a lattice homomorphism satisfying the conditions of Definition 4.2.3, provided the lattice homomorphism is not a constant mapping.*

Proof: We give a trivial counterexample. Consider a bilattice $\mathbf{D} = \langle \{a, b\}, \leq_1, \leq_2 \rangle$ where the order relations are equal and it holds: $a \leq_1 b$ and $a \leq_2 b$, but also $b \not\leq_1 a$ and $b \not\leq_2 a$. Obviously, $\mathbf{D}$ is a bilattice and in fact, it is an interlaced bilattice. It is clearly impossible to introduce a homomorphism $\neg : \{a, b\} \longrightarrow \{a, b\}$, such that $a \leq_1 b \Rightarrow \neg a \leq_1 \neg b$ and also $a \leq_2 b \Rightarrow \neg b \leq_2 \neg a$, if one does not allow the constant mapping $\neg : \{a, b\} \longrightarrow \{a\}$ or the dual mapping $\neg : \{a, b\} \longrightarrow \{b\}$. This suffices to show the fact. q.e.d.

Notice that a constant mapping $\neg$ always satisfies the lattice homomorphism condition. The next proposition specifies some important properties of $\neg$ in interlaced bilattices that was originally proven by Melvin Fitting.[4]

Proposition 5.4.2 *Assume* $\mathbf{D} = \langle D, \leq_1, \leq_2, \neg \rangle$ *is an interlaced bilattice with a lattice homomorphism* $\neg$ *(negation). Then, for every subset* $A \subseteq D$ *the following equalities holds:*

(i) $\neg \bigwedge A = \bigwedge \neg A$
(ii) $\neg \bigvee A = \bigvee \neg A$
(iii) $\neg \Pi A = \Sigma \neg A$
(iv) $\neg \Sigma A = \Pi \neg A$

Proof: We show claim (i). The other claims (ii) - (iv) are dual claims of (i). Assume $A \subseteq D$ is arbitrarily given. Then, it holds for all $a \in A : \bigwedge A \leq_1 a$. Because of the definition of $\neg$, we have $\neg \bigwedge A \leq_1 \neg a$ for all $a \in A$. Hence: $\neg \bigwedge A \leq_1 \bigwedge \neg A$. This shows one direction of the claim. Concerning the other direction, for all $a \in A$ it holds: $\bigwedge \neg A \leq_1 \neg a$ by the definition of $\bigwedge$. Applying $\neg$ yields the relation $\neg \bigwedge \neg A \leq_1 a$ (for all $a \in A$). Hence, $\neg \bigwedge \neg A \leq_1 \bigwedge A$ and therefore, $\neg\neg \bigwedge \neg A \leq_1 \neg \bigwedge A$ which is equivalent to $\bigwedge \neg A \leq_1 \neg \bigwedge A$. This shows the other direction of the claim. q.e.d.

Remark 5.4.1 (i) Notice that the above proposition is a generalization of the DeMorgan rules: For $A = \{a, b\}$ we have the well-known equality: $\neg(a \cdot b) = \neg a + \neg b$. In other words, with respect to the truth order relation ($\leq_2$ order relation) the behavior of $\neg$ has the same properties as the ordinary negation of classical logic.

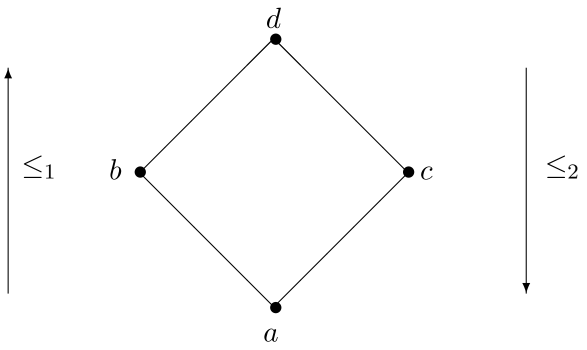
(ii) There is a second possibility to introduce a negation, namely the negation $\sim$ that preserves the $\leq_2$ order and inverts the $\leq_1$ order. More precisely, we require that $\sim$ satisfies the following two conditions (for a given bilattice $\langle D, \leq_1, \leq_2, \sim \rangle = \langle D, \wedge, \vee, \cdot, +, \sim \rangle$):

$\sim$: $\langle D, \wedge, \vee \rangle \longrightarrow \langle D, \vee, \wedge \rangle$
$\sim$: $\langle D, \cdot, + \rangle \longrightarrow \langle D, \cdot, + \rangle$

The operation $\sim$ can be understood as the dual operation of $\neg$. A dual form of Proposition 5.4.2 holds also for the dual negation $\sim$. This is justified by the fact that bilattices are completely symmetric. An interesting question concerns the relations between the lattice homomorphisms $\neg$ and $\sim$ in an interlaced bilattice that contains both negations. We neither examine these relations, nor possible applications for the new negation $\sim$, because we will not need $\sim$ in the further examination of Kripke's fixed point approach. Therefore, we skip a discussion of the relationship between the two negations $\neg$ and $\sim$ here.

[4] Cf. [Fi89].

In order to get a result that tells us which types of bilattices allow to introduce a negation in the sense of Definition 4.2.3, it is necessary to assume that the order relations of the considered bilattice are distinct ($\leq_1 \neq \leq_2$), and that $\neg$ is not a constant function. That are immediate consequences of the above consideration. It turns out that the requirement that the order relations are distinct is not sufficient for the possibility to introduce a negation $\neg$, either. Notice that the following bilattice is interlaced, but nevertheless does not allow the introduction of a lattice homomorphism $\neg$ in the sense of Definition 4.2.3.



The problem is that $\leq_2$ is simply the dual order of $\leq_1$. A last point concerns the bilinearity property of bilattices: If the bilattice is not bilinear, then it is not possible to introduce a non-trivial lattice homomorphism in general. In order to see this, consider the bilattice in Remark 5.1.2(v). Therefore, we can conclude that the described conditions are necessary conditions for the possibility to introduce a non-trivial lattice homomorphism in the sense of Definition 4.2.3 on an interlaced bilattice $\langle D, \leq_1, \leq_2 \rangle$. We summarize the insights up to now. If $\langle D, \leq_1, \leq_2 \rangle$ is an interlaced (non-trivial) bilattice and it is possible to introduce a non-trivial negation $\neg$, then the following must hold:

- The order relations $\leq_1$ and $\leq_2$ are not equal.
- The order relation $\leq_2$ is not the dual of order relation $\leq_1$.
- $\neg$ is not a constant function.
- The bilattice is bilinear.

What can be said about sufficient conditions to introduce a lattice homomorphism? The next Proposition 5.4.3 specifies the finite case where the bilattice is generated by the product of two chains.

Proposition 5.4.3 *Assume $\langle \{a_1, a_2, ..., a_n\}, \leq \rangle$ and $\langle \{b_1, b_2, ..., b_n\}, \leq' \rangle$ are two chains. Construct the product bilattice $\langle D, \leq_1, \leq_2 \rangle$ generated by these chains. Then, a negation $\neg$ can be defined on this bilattce.*

Proof: Assume the premises of the proposition. We apply the standard coding of the domain $\{a_1, a_2, ..., a_n\} \times \{b_1, b_2, ..., b_n\}$. First, notice that the product bilattice is interlaced[5]. Then, we can construct the negation $\neg$ as follows. $\neg(\langle a_i, b_i \rangle) = \langle a_i, b_i \rangle$ for all $i \in \{1, 2, ..., n\}$. For all pairs $\langle a_i, b_j \rangle$ for $i \neq j$ we define: $\neg(\langle a_i, b_j \rangle) = \langle b_j, a_i \rangle$. It is obvious that this definition of $\neg$ suffices to show the claim. q.e.d.

The above proposition shows how we can introduce a negation $\neg$ into a bilattice that is completely symmetric and a (finite) product of chains. What happens, if the two chains do not have the same length? The following fact shows that in this case a negation cannot be introduced.

Fact 5.4.4 *Assume a product bilattice $\langle D, \leq_1, \leq_2 \rangle$ is generated by two finite chains $\langle \{a_1, a_2, ..., a_n\}, \leq \rangle$ and $\langle \{b_1, b_2, ..., b_m\}, \leq' \rangle$ where $n \neq m$. Then, a negation $\neg$ satisfying the conditions of Definition 4.2.3 cannot be introduced.*

Proof: It is easy to see that even the product bilattice generated by the two chains $\langle \{a_1, a_2\}, \leq \rangle$ and $\langle \{b_1, b_2, b_3\}, \leq' \rangle$ suffices to show that a negation $\neg$ cannot be introduced. One has simply to check some cases. An easy induction shows that this is also true for arbitrary $n, m \in \omega$ with $n \neq m$. q.e.d.

Using Proposition 5.4.3 and Fact 5.4.4, we can generalize the product approach to infinite chains.

Fact 5.4.5 *Assume a bilattice $\langle D, \leq_1, \leq_2 \rangle$ is generated by two infinite chains $\langle \{a_1, a_2, ..., a_\alpha\}, \leq \rangle$ and $\langle \{b_1, b_2, ..., b_\alpha\}, \leq' \rangle$ where α is an ordinal. Then, a negation $\neg$ can be introduced.*

Proof: Applying the same reasoning as in Proposition 5.4.3 using ordinals instead of natural numbers yields the desired result. q.e.d.

So far we have seen necessary conditions for the introduction of a negation $\neg$ in a given bilattice and sufficient conditions for such an introduction. A characterization result would give us criteria that are necessary and sufficient for the possibility of the introduction of $\neg$. Notice that it is not a necessary condition for a bilattice to be a product of chains in order to make a negation possible. In Example 4.2.3(iii) we saw a bilattice that is not a product of chains but neverthelss is a bilattice with negation $\neg$. The symmetry condition on bilattices is significant. We saw in Fact 5.4.4 that asymmetric bilattices do not work in order to introduce a negation. Roughly speaking, the two main conditions that guarantee the introduction of a negation are bilinearity and a certain symmetry condition. This symmetry condition is induced by the properties of $\neg$ inverting the $\leq_2$ order relation and preserving the $\leq_1$ order relation. The following definition makes the concept of a symmetric bilattice precise.

[5] Cf. Fact 5.1.3(v)

Definition 5.4.6 *Assume* $\mathbf{D} = \langle D, \leq_1, \leq_2 \rangle$ *is a bilattice where* $\leq_1 \neq \leq_2$*, and* $\leq_1$ *is not the complement relation of* $\leq_2$*.* $\mathbf{D}$ *is called symmetric if and only* $\langle D, \leq_1, \leq_2' \rangle$ *is isomorphic to* $\mathbf{D}$ *where* $\leq_2'$ *is the inverse order relation of* $\leq_2$*.*

Proposition 5.4.7 *Assume* $\mathbf{D} = \langle D, \leq_1, \leq_2 \rangle$ *is a bilinear bilattice, such that* $\leq_1 \neq \leq_2$*, and* $\leq_1$ *is not the dual of* $\leq_2$*. A non-constant lattice homomorphism satisfying the conditions of Definition 4.2.3 can be defined on* $\mathbf{D}$ *if and only if the bilattice is symmetric.*

Proof: "$\Rightarrow$" Assume $\mathbf{D}$ is a bilinear bilattice, such that $\leq_1 \neq \leq_2$, and $\leq_1$ is not the dual of $\leq_2$. Assume that $\mathbf{D}$ is not symmetric. Then, there is no isomorphism f mapping $\mathbf{D}$ to the bilattice $\langle D, \leq_1, \leq_2' \rangle$. Hence, there is no homomorphism f mapping $\mathbf{D}$ to $\langle D, \leq_1, \leq_2' \rangle$. Therefore, there cannot be a homomorphism $\neg$ that satisfies the conditions of Definition 4.2.3.
"$\Leftarrow$" Assume $\mathbf{D}$ is symmetric and bilinear. Then, the isomorphism $f : D \longrightarrow D$ induced by the symmetry condition maps every element $d \in D$ to an element $d' \in D$. Define the negation $\neg$ by $\neg d = d'$ for all $d \in D$. It is easy to check that $\neg$ satisfies the conditions of Definition 4.2.3. q.e.d.

Remark 5.4.2 (i) Notice that the interlacing condition is not a crucial condition for the possibility to define a lattice homomorphism $\neg$. There are many examples of bilattices that are not interlaced and where a lattice homomorphism can be defined (compare for instance Example 4.2.3(iii)). In the following, we shall restrict our attention to bilattices that are interlaced, because these are the important structures we are dealing with.

(ii) The characterization of Proposition 5.4.7 shows that bilattices with a negation $\neg$ are quite regular structures. Notice that it is not true that the isomorphism f in symmetric bilattices induces a certain kind of complement operation. It is not necessary that the supremum with respect to $\leq_1$ of a point $d \in D$ and the dual point $f(d)$ is the top element of $\leq_1$. The same holds with respect to the infimum and the order relation $\leq_2$.

In the next section, we shall consider certain applications of the characterization theorems proven in this section where we focus on applications in the context of Kripke's construction.

5.5 Some Applications

The characterization results in the above subsection make clear that the class of interlaced bilattices contains subclasses of bilattices with interesting properties. In the following, we will give some applications of these characterization results in the context of Krikpe's theory of partially defined truth predicates.

Assume a language L is given that includes ordinary arithmetic (or from a more abstract perspective a so-called elementary coding scheme in the sense of [Mo74]). Assume further that we work in positive logic, i.e. in a logic that does not contain an odd number of negation symbols in front of any expression in L. Notice that as a consequence of this requirement no implication of the form $p \rightarrow q$ can be formulated in this logic where p and q are arbitrary formulas of the given language. This holds, because the implication $p \rightarrow q$ is equivalent to $\neg p \vee q$ and the latter formula contains an odd number of negations. The possible assignments for formulas are not restricted, i.e. any formula can be evaluated as true, false, neither true nor false, or both true and false.

It is clear that the Liar sentence cannot be represented in this logic, because a negation is essentially needed to represent the Liar sentence. Notice further that a Truth-teller sentence can be formulated, because we have a coding scheme that allows us to build pathological sentences not containing a negation symbol. The Truth-teller is one example of such a sentence. We will examine the behavior of the Truth-teller more closely now.

Assume a Kripke-style construction based on L and a four-valued positive logic are given.[6] Which truth value do the maximal and minimal fixed points of this construction assign to the Truth-teller sentence? In the minimal fixed point w.r.t. the information order $\leq_I$, the Truth-teller is neither true nor false. This holds, because the minimal fixed point codes the minimal amount of information concerning this sentence. This means that the Truth-teller is neither true not false, because this assignment represents the underspecified situation. Clearly, in the maximal fixed point of the $\leq_I$ order this very sentence is both true and false, because this extension remains fixed in every further step of the construction and encodes the maximal (in our case inconsistent) information concerning the Truth-teller sentence.

What happens in the minimal and maximal fixed point w.r.t. the truth order relation $\leq_T$? Here we have a quite similar picture: in the minimal fixed point the Truth-teller is false (minimal w.r.t. the truth order $\leq_T$) and in the maximal fixed point this sentence is true (maximal w.r.t. the truth order $\leq_T$). This is an immediate consequence of the biconsistency of the Truth-teller. Together we get all four possible assignments.

Consider now our first characterization result (Theorem 5.1.10). Because of this theorem the following four conditions necessarily hold:

For all nonempty $A \subseteq D : \bigwedge A = \Pi A \wedge \Sigma A$
For all nonempty $A \subseteq D : \Pi A = \bigwedge A \cdot \bigvee A$
For all nonempty $A \subseteq D : \bigvee A = \Pi A \vee \Sigma A$
For all nonempty $A \subseteq D : \Sigma A = \bigwedge A + \bigvee A$

The truth value of the Truth-teller w.r.t. to the information order can be calculated by considering the truth values of this sentence in the minimal and maximal fixed points w.r.t. the truth order. Because this sentence is false in the minimal fixed point and true in the maximal fixed point with respect to the

[6] We assume that the logic corresponds algebraically to the interlaced bilattice `FOUR`.

truth order, the minimal truth value with respect to the information order can be calculated by $F \wedge T = N$. Similar relations hold with respect to the other three combinatorial permutations of the order relations. For example, in order to calculate the truth value of the maximal fixed point w.r.t the truth order, we simply perform the following calculation: $N + B = T$.

Notice that we are not forced to restrict out attention to the minimal and maximal fixed points of the whole domain of all fixed point. Every nonempty subset of fixed points behave in a similar way and the value of the Truth-teller sentence can be calculated by the truth values of the infimum and the supremum of this subset with respect to the other order relations. The calculation of the extremal points of a given set can be interpreted as a closure. Assume for example, we examine a set of points that contains only assignments of sentences that are true, false, or neither true nor false. Calculating the maximal point with respect to the information order results (in general) in a point where the Truth-teller sentence will be assigned the truth value both true and false. Roughly speaking, there is too much information contained in the subset. In order to make the set consistent in a four-valued logic, the Truth-teller gets the assignment B in the maximal fixed point of the information order.

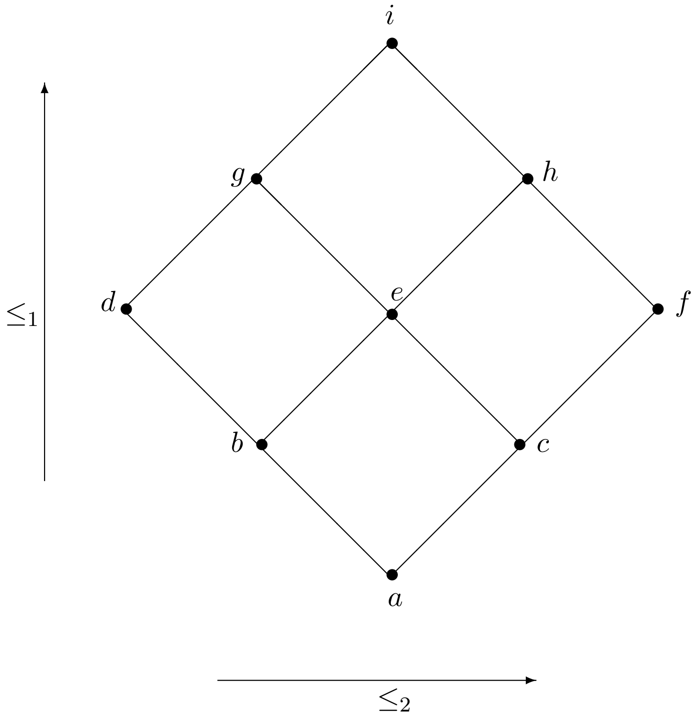
Calculations of this kind can be performed quite generally in interlaced bilattice. Not only single sentences like the Truth-teller can be examined by the characterization result Theorem 5.1.10, but collections (sets) of sentences (pathological ones as well as non-pathological ones). For example, consider a subset of fixed points $X \subseteq \{T, F, N, B\}^{Sent_{L^+}}$ consisting of two points: $X = \{x, y\}$. Assume further that we are not interested in the evaluation of all sentences in a particular fixed point but only in certain 'interesting' sentences (for example pathological ones or sentences that were differently evaluated by the experts in the beginning). Given X we know the minimal and the maximal fixed point with respect to the $\leq_T$ order relation. Then, these two fixed points are completely sufficient to calculate the minimal fixed point in the $\leq_I$ order relation. That means, we can calculate the consensus fixed point where no inconsistent information arises as well as the maximal fixed point with respect to the information order where we accept every available information. With the technical tools in the background it is easy to figure out where the differences between the experts is located and what consequences this has for the other sentences.

In total, we can say that first, there is the possibility to calculate the truth value for extremal points of single sentences provided the truth values of these points w.r.t. the alternative order relation are known. Second, we can calculate certain fixed points for a given set of sentences provided we have sufficient information of the behavior of these sentences in other fixed points.

The theorems in this chapter are completely general. They work for the class of interlaced bilattices and the specified subclasses of interlaced bilattices on an abstract level. An important example for an interlaced bilattices used in order to model presuppositions in natural language is the bilattice `NINE`.[7] The

[7] Cf. [Sc96].

following diagram specifies the order theoretic structure of NINE.



This bilattice was extensively used in [Sc96] in order to develop a theory of presuppositions. With an appropriate interpretation of the truth values it is easy to mirror Kripke's construction in this bilattice. The truth values are usually given by pairs of numbers:

$a = \langle 0, 0\rangle$
$b = \langle 1/2, 0\rangle$
$c = \langle 0, 1/2\rangle$
$d = \langle 1, 0\rangle$
$e = \langle 1/2, 1/2\rangle$
$f = \langle 0, 1\rangle$
$g = \langle 1, 1/2\rangle$
$h = \langle 1/2, 1\rangle$
$i = \langle 1, 1\rangle$

The pairs can intuitively be interpreted as specifying the degree of falseness and truthfulness. For example, $\langle 0, 1\rangle$ can simply be interpreted as true whereas $\langle 1/2, 1/2\rangle$ can be interpreted as both slightly false and slightly true. The $\leq_1$ order relation can still be associated with an increase of information whereas the $\leq_2$ relation can be identified with an increase of truthfulness and also a decrease of falseness. From this perspective the general treatment is similar to the one for the logics considered so far.

In NINE, more possibilities arise concerning the behavior of the truth values of pathological sentences. For example, the Truth-teller can range over all nine truth values whereas the Liar sentence can neither get the definite truth value $\langle 1,0\rangle$ nor $\langle 0,1\rangle$.[8] It is possible to assign the truth values $\langle 0,0\rangle$, $\langle 1/2,1/2\rangle$, and $\langle 1,1\rangle$ to the Liar sentence (relative to appropriate fixed points). Hence, there is a wider range of possible interpretations in NINE in comparison with the variety in FOUR. In general, the calculations of maximal and minimal fixed points remain similar to the presented ideas in this chapter.

From a complexity theoretic standpoint Theorem 5.1.10 is quite interesting. Whereas the original definition of the interlacing property of bilattices is quite complex, the first characterization theorem reduces this complexity significantly. Assume $\mathbf{D} = \langle D, \leq_1, \leq_2\rangle$ is a bilattice, such that $|D| = n$. In order to check the interlacing property, one has to perform $2^n \cdot 2^n \cdot 4 = 2^{2n+2}$ many calculations.[9] Using the characterization Theorem 5.1.10 of interlaced bilattices the total number of calculations reduces to $2^n \cdot 4 = 2^{n+2}$. From a computational perspective (at least for finite bilattices) this reduction in the complexity is significant.

If one wants to show that a certain bilinear bilattice is interlaced, one can check the distributivity condition according to Theorem 5.3.1. For example, assume a bilinear bilattice $\langle D, \leq_1, \leq_2\rangle$ with $|D| = n$ is given. In order to check whether $\langle D, \leq_1, \leq_2\rangle$ is interlaced or not one has to perform $12 \cdot 2 \cdot n^3$ many calculations using the distributivity condition. In comparison to that, 2^{2n+2} many calculations are necessary in order to check the interlacing condition if we apply the ordinary definition. Here, the complexity reduces from an exponential behavior to a cubic one.

We finish that chapter with these remarks. In the next chapter we will discuss problems of Kripke's theory of truth as well as.

5.6 History

Bilattices were introduced and originally examined in [Gi86] and [Gi88]. Motivated by the question how bilattices can be used in order to model Kripke's theory of truth in a generalized context, Fitting developed the concept of an interlaced bilattice. He also uses Ginsberg's notion of a distributive bilattice in [Fi89]. Although one of Fittings applications of the theory of bilattices is the theory of truth, he examined other applications, too, most prominently in the theory of programming languages and theoretical computer science in general

[8] We assume that an appropriate interpretation of the logical connectives is given. For more information the reader is referred to [Sc96].

[9] There are 2^n many subsets of D, and we have to consider all pairs of subsets. Therefore, we have $2^n \cdot 2^n$ many pairs of subsets of D. Four calculations of the form $A \leq B \rightarrow sup(A) \leq sup(B)$ must be checked. This determines the total number of necessary calculations.

(compare [Fi88, Fi91, Fi93] and [Fi94]). In [ArAv96], bifilters, prime bifilters, and ultrabifilters were introduced in order to prove a completeness result for so-called bilattice logic. Some of these ideas were developed further in [Sc96] and applied to a theory of presuppositions. All the theorems and characterizations in Section 5.1 to 5.3 are based on [Ku99]. The characterization theorem 5.3.1 is explicitly proven in this paper. Although it seems to be a natural question under which conditions a lattice homomorphism interpreted as a negation can be introduced in an interlaced bilattice (as well as in other types of bilattices) the author does not know of any other work concerning this point. The idea to introduce different kinds of negations (in our context the negations $\neg$ and $\sim$) goes back to Fitting and Ginsberg (compare [Fi89] and [Gi88]).

Chapter 6

Problems of Kripke's Approach

In this chapter, we shall examine various forms of criticisms that were mentioned by authors in a more or less direct reaction to Kripke's fixed point approach. It is not possible to discuss everything that was published concerning critical aspects of Kripke's construction, because the amount of literature is enormous and the number of papers and books dealing with problems of Kripke's account is so large that it is hard to keep track of the topic.[1] We will stress certain crucial points that seem to be very important for further developments concerning the treatment of truth predicates in natural languages. We do not claim to give a complete overview of the topic in any respect.

First, we shall classify different types of criticism in groups, in order to be able to get an overview of the situation. We choose four major groups in which different kinds of criticisms can be subsumed. The first group contains empirical facts that make clear that certain discourse patterns involving the truth predicate cannot be modeled in Kripke's account, or at least cannot be modeled in the form Kripke proposed originally. This is a form of criticism Kripke uses himself with respect to Tarski's typed languages: Kripke shows in [Kr75] that Tarski's account cannot provide a correct modeling of certain discourse situations. This type of criticism can be interpreted as pointing out a defect of the empirical situation we are confronted with when applying the fixed point approach.

The second group of criticism claims that - even though he tries to avoid typed languages explicitly - Kripke's account nevertheless implies a very similar typing of languages as in Tarski's approach (at least for the modeling of certain examples). This is a problem that was seen by Kripke himself. Unfortunately, he never proposed a solution for this problem.

Third, it is clear that using one or the other kind of a non-classical logic (in our case a four-valued logic, in Kripke's original account a three-valued logic) provokes opposition of logical purists. People who claim that logic is essentially

[1] A good overview of comments, criticism, and ideas for a further development of the fixed point approach of Saul Kripke is [Ma84].

classical logic, and logical truths are a priori truths[2] do have problems to except the modeling of phenomena that uses non-classical reasoning. Logical pluralism (and that position must be assumed in the case one argues for Kripke's ideas[3]) is in philosophical considerations not generally accepted. This group of criticism is summarized under the label *conceptual problems* and is the third group of criticism we will examine.

Last but not least, we want to mention a fourth group of criticism. This group deals with questions of the ontological nature of the fixed point approach. Is it intuitively correct that one uses an infinite process in order to define an infinite number of (possible) truth predicates, and after the production of these (possible) truth predicates one is forced to choose one for the (correct) truth predicate of natural language? Do we really need infinitely many truth predicates? Another aspect in this group is the status of the considered objects in question: Kripke argues for the position that sentences are the bearers of truth, whereas propositions do not play any role in Kripke's considerations. The question which object can be the bearer of truth is philosophically quite important. We will begin our discussion with an examination of some empirical problems.

6.1 Empirical Problems

For a large number of examples, Kripke's treatment of pathological sentences models quite nicely the intuition of native speakers. Additionally, this account enables us to give a definition of the truth predicate on the object language level. But major problems can arise if one tries to model special kinds of short discourses and ordinary human reasoning. Although Kripke did not develop his fixed point approach in order to model discourses (and therefore he is in a position to reject criticisms of that kind by pointing out that the modeling of discourses was not his intention) one has to question whether Kripke's ideas can be extended to deal with this challenge. The crucial point is that there are principal reasons for the impossibility of enlarging Kripke's fixed point approach to get a correct representation of these examples. Only if such an extension is possible the fixed point account can count as appropriate for the modeling of the truth predicate of natural language. That is quite clear if one considers newer results in linguistics and in particular the struggle to find the correct semantics for discourse representation. It is highly questionable whether discourses can be properly modeled in fixed point accounts, because for Kripke there is not even space to model propositions not to mention the modeling of contexts in general. He deals exclusively with sentences. We will discuss this challenge more closely in the following.

Let us consider the Gupta puzzle again.[4] We mentioned this example in

[2] Compare [Bo97] for a very new version of a philosophical position that establishes laws of classical logic as a priori truths.

[3] Feferman (as well as Aczel) proposed a Kripke-style construction using classical logic (cf. [Fe84] and [Ac78]). These constructions are based on a partial model theory. The claim that one is forced to apply a non-classical logic is nevertheless true if one takes Kripke literally.

[4] Cf. [Gu82].

Chapter 2 already (cf. 2(7)). The Gupta puzzle features a very interesting kind of reasoning: in order to evaluate the truth values of sentences uttered by a person we have to refer to sentences uttered by another person. In order to evaluate the utterances of the second person, it is necessary to know the truth values of the sentences of the first person. The interesting point is that in ordinary common sense reasoning, the problem has a unique solution, whereas in Kripke's fixed point model the sentences are assigned altogether the truth value N, therefore cannot be properly evaluated. We formulate the example once more in a slightly simplified form.

Example 6.1.1 Consider the following situation: Two individuals C and D play cards in which D has the ace of clubs (background information). Now the following statements uttered by A and B - two observers of the card game - takes place.

A claims:	(a1) C has the ace of clubs.
	(a2) All claims made by B are true.
	(a3) At least one of the claims made by B is false.
B claims:	(b1) D has the ace of clubs.
	(b2) At most one of the claims made by A is true.

Now we can ask the question, whether it is possible to assign unique truth values to the statements made by A and B using common sense reasoning (in a two-valued logic). Using common sense reasoning, it is clear that there is a solution to this problem. The statements (a2) and (a3) contradict each other and cannot be true at the same time. Therefore, at most one of these two claims must be true and one must be false[5]. Because of that reasoning, we know that (b2) is true. Furthermore, the background knowledge guarantees that statement (a1) is false. This solves the problem because by background information we know that (b1) is true, and finally we can conclude that (a2) must also be true. Notice that this is the only consistent analysis of truth value assignments of the Gupta puzzle, provided one works in a classical two-valued logic and one has the background information that D has the ace of clubs. Therefore, from an intuitive (naive) point of view this puzzle is not difficult to solve because it is plainly an exercise in (classical) logical reasoning, checking cases and finding the right solutions, although the statements of A and B refer to each other.

Taking into account that the above reasoning is not difficult at all, it is even more surprising that in Kripke's quite sophisticated theory of truth it is not possible to represent the above reasoning. What is the problem if one tries to evaluate the puzzle in Kripke's theory? Difficulties arise if one wants to assign truth values to (a2) and (a3). In order to assign truth values to these sentences, it is necessary to know the truth values of (b1) and (b2). Concerning

[5] At this point it is crucial to work in a classical logic, because the bivalence principle must hold.

(b2), this is only possible if we know the truth values of all sentences uttered by A in particular the truth values of (a2) and (a3). Notice that in Kripke's account, at every stage the truth value of (a2), (a3), and (b2) will remain underdetermined (therefore these statements are assigned the truth value N). We can conclude that in Kripke's account it is not possible to assign definite truth values to all sentences of Example 6.1.1. This result is highly counterintuitive, if one considers the solution above using simply common sense reasoning.[6]

Another problem in Kripke's account is caused by the fact that complex sentences that include the Liar sentence in one or the other form (e.g. complex sentences of the form $\phi(\lambda)$ where λ is (the code of) the Liar sentence) sometimes are assigned counterintuitive truth values. Consider, for example, the following sentence (1) where ϕ is an arbitrary sentence of a given language L, not necessarily a pathological one.

(1) $\phi \vee \neg\phi$

(1) is a classical tautology (in classical logic). Usually, we want that a semantics for natural language preserves classical tautologies like (1). Therefore, it seems intuitively right to assign the truth value T to (1). Unfortunately, this is not the case for every substitution instance in (1) in Kripke's fixed point approach. Consider, for example, the Liar sentence λ and substitute λ for ϕ in (1). Obviously, the newly created sentence $\lambda \vee \neg\lambda$ does not have a definite truth value at any stage of the fixed point construction, because the constituents (λ and $\neg\lambda$) cannot be evaluated on any level of this construction.[7] The consequence, namely that (1) gets the same truth value as the Liar sentence, seems to be counterintuitive. The crucial reason for that fact can be found in the properties of three- (or four-) valued logic: both do not preserve classical tautologies.

Another very similar example for this phenomenon, namely that Kripke's account causes counterintuitive evaluations if one states classical tautologies are sentences like (2).

(2) $\forall\phi \in L : \neg(\mathbf{T}(\phi) \wedge \neg\mathbf{T}(\phi))$

In (2), $\mathbf{T}$ represents a Kripke style truth predicate, for example, the maximal intrinsic fixed point (or the minimal fixed point) if one works in a three-valued logic, or the minimal fixed point if one works in a four-valued logic of Kripke's construction. In four-valued logic, it is clear that sentences that are assigned the truth value B (true and false) are constitutive for this logic. Therefore, in such a logic it is obvious that (2) cannot be true in general. Let us restrict our

[6] Examples like 6.1.1 were a strong motivation for Anil Gupta to develop his revision theory of truth (cf. [Gu82]).

[7] Notice that this is independent of the choice of the particular non-classical logic (or algebraic structure). Sufficient for the existence of this problem is a logic that does not preserve classical tautologies like the examined three-valued or four-valued logics we examined in this part of the work.

attention therefore to the three-valued case. Working in a three-valued logic, and applying Kripke's fixed point approach, (2) would be neither true nor false at every stage. That holds, because in order to evaluate (2), one has to know the truth values of all sentences in the language. For the Liar sentence λ there is no definite truth value on any stage, therefore it is not possible to assign a definite truth value to (2).

Even - and that is much more important - if there does not exist a Liar-like sentence at all in the language (for example, assume that the language has no negation or there is no coding scheme that allows us to construct self-referential sentences), even then (2) cannot be assigned a definite truth value. That is true because in order to evaluate (2), one has to know all truth values of the other sentences, in particular one has to know the truth value of (2) itself. And here is the crucial point where the circularity (on a meta-level) comes in. It is impossible to assign a definite truth value to (2) as long as (2) itself is not evaluated. In order to evaluate (2), one needs the evaluation of (2) itself. We are in a vicious circle when we apply Kripke's fixed point approach.

Gupta criticizes these properties of Kripke's theory because of the counterintuitive consequences.[8] Although Gupta mentions crucial points, it should not be surprising that this behavior happens to be the case. The considered three-valued logics (based on the strong Kleene-tables) do not have tautologies at all (even a classical tautology like $p \rightarrow p$ is not a tautology in the proposed three- as well as four-valued logic). Then, nobody can reasonably require that a certain particular sentence like (2) should be a tautology. In the four-valued case, the same is true because the three-valued case is only a specialization of the four-value case. Here, there is a strong relation to the group of criticisms of Kripke's account that reject the usage of non-classical logic.[9] A strict separation of these groups of criticism is not possible in this context.

For someone who does not believe in logical pluralism, I think that Gupta's point is a striking argument against Kripke. Logical pluralists are more relaxed in such situations. If there are no tautologies any longer in our logic, then it would be rather strange to hope to get tautologies in our semantics that is based on that logic. In that respect, one cannot hope to get a feature that is explicitly excluded from the very beginning.

There is a further classical problem of Kripke's account that is strongly connected with the hidden typing of languages in the fixed point approach.[10] Consider the strengthened Liar sentence.

(3) The Liar sentence is not true.

One can show that even in Kripke's fixed point approach one can deduce a contradiction from (3). We sketch the reasoning that yields a contradiction.

[8] Cf. [Gu82].

[9] Compare Subsection 6.4.1 for more information concerning the usage of a non-classical logic.

[10] Compare Section 6.2 for more information concerning this point.

> Assume (3) is true, then it follows that (3) is not true.
> First possibility: If (3) is false, then sentence (3) is true.
> Second possibility: If (3) is neither true nor false, then (3) is true.
> Conclude: If (3) is true, then (3) is true and not true and if (3) is not true, then (3) is true and not true. In other words, we have a contradiction.[11]

The problem is that one can shift the difficulties of the Liar sentence to the three-valued (or four-valued) account. The assumption we need in order to derive the contradiction is that there is 'trivalence' (or 'quadrivalence') principle.[12] Kripke was aware of this problem and - strangely enough - he proposed to introduce levels (as Tarski did) in order to fix this problem. Hence, Kripke does not avoid a major problem he figured out as a problem in Tarski's account. In the next section, we will consider this problem of hidden types of truth predicates more closely.

6.2 The Hidden Types of Languages

The second category of criticism we mentioned in the introduction of this chapter shows some similarities to the fact that the strengthened Liar sentence cannot be modeled in Kripke's fixed point approach. Authors supporting this criticism claim that Kripke's real goal, namely to improve Tarski's account to avoid typed languages, has failed. Obviously, there is the possibility to define a truth predicate in the object language using Kripke's fixed point construction. Hence, on the surface, Kripke is an improvement in comparison with Tarski, because he avoids (at first sight) hierarchies of languages (as was crucial in Tarski's account). Nevertheless Kripke cannot get rid of hierarchies of typed languages. In order to see the problem clearly, we develop in this section a more sophisticated argument of the existence of hidden typed languages in Kipke's account.

Assume a language L is given, and assume further that L is strong enough to code Liar-like sentences. Now, consider again example (3). What is a reasonable interpretation of (3)? As we know from Section 6.1, the strengthened Liar sentence is a paradoxical sentence that yields a contradiction in Kripke's fixed point approach. Kripke's fixed point approach cannot assign the third truth value to (3), because this sentence essentially changes its truth value in every new level as the ordinary Liar sentence does in ordinary classical logic.

First, we have to conclude that (3) is another example of a sentence that is assigned a counterintuitive truth value in Kripke's fixed point construction. Second, we can speculate that there is a hidden implicit typing of the languages. Sentences like (3) (i.e. circular sentences that include a statement about a truth value different from T and F) could be considered as sentences of another

[11] Notice that the ordinary Liar sentence does not have this behavior: the truth value N (or B in the four-value account) guarantees that the Liar sentence is stably N (or B).

[12] Clearly, any finite number n of truth values would not solve this problem, because the problem of the strengthened Liar can be shifted to any n-valence logic.

type. Then, we can find an elegant way out of the problems.[13] The price we have to pay is that we are forced to accept typed languages and truth predicates of different kinds, although it was one main motivation of fixed point approaches to avoid stages of truth predicates. Although Kripke can model several examples intuitively correctly, there is the possibility to shift Tarski's problem to another level where again a hierarchy of languages is needed in order to avoid inconsistencies of the interpretation.

It is worth it to mention that Kripke himself recognized that problem as the following quotation shows:

> *"If we think of the minimal fixed point, say under the Kleene valuation, as giving a model of natural language, then the sense in which we can say, in natural language, that a liar sentence is not true must be thought of as associated with some later stage in the development of natural language, one in which speakers reflect on the generation process leading to the minimal fixed point. It is not itself a part of that process. The necessity to ascend to a metalanguage may be one of the weaknesses of the present theory. The ghost of the Tarski hierarchy is still with us."*[14]

How should we interpret this situation? The ghost of Tarski's hierarchy forces us to use different truth predicates, if we want to model natural languages. One truth predicate can be used when we deal with Liar-like sentences, but another truth predicate is needed when we try to model a particular reasoning about the Liar sentence itself (and similar constructions). It is clear that this brings us far away from an analysis in which we can define a truth predicate that corresponds intuitively to the truth predicate in natural language. The idea that there is no unique truth predicate in natural language but infinitely many different ones, is for most researchers a non-acceptable position. Furthermore, it is counterintuitive to develop a sophisticated theory in order to avoid a hierarchy of languages but nevertheless accepts a hierarchy of languages on another level.[15] Kripke's ideas are therefore good advice to avoid counterintuitive consequences in his theory, but in order to model natural language as something else than his fixed point approach is essentially needed.

Anil Gupta (in [GuBe93]) strengthens the hierarchy problem by claiming that the problem of modeling the truth predicate of a given language in a descriptively correct way cannot be achieved even by fixed point approaches because of the hidden types of truth predicates. Gupta's claim is that hierarchies of truth predicates destroy the descriptive appropriateness of the truth theory in question. We consider the following two sentences.

[13] Notice that this strategy would weaken Gupta's criticism concerning the counterintuitive example (1) and (2).

[14] Cf. [Kr75], p.80.

[15] The problem is that fixed point theories do not allow full reflection principles. A deeper reason for this is the fact that they are well-founded and inductive theories and hence cannot be applied to themselves.

(4)(a) The pope knows all true sentences.
(4)(b) The pope does not know all true sentences.

Although it is possible to model (4)(a) in fixed point theories by introducing indices for the truth predicates, it is (according to Gupta) completely unclear what (4)(b) should mean if one works in a theory of hierarchies. The problem is that it is at least as complicated to develop a theory of levels (types) as to develop a theory of truth, because the theory of levels has to be also a theory of itself and the same problem of circularity arises. Again the deeper reason for this behavior is the fact that fixed point theories are essentially inductive and therefore well-founded theories. A full reflection principle (that can be applied to the theory itself) cannot be established in a fixed point approach. Even here the real problem of a theory of truth is shifted to another metalevel.

The vicious circle we are automatically involved when modeling the truth predicate of natural language brings us back to the problems we started with. In the next section, we will consider some conceptual problems of Kripke's fixed point approach.

6.3 Conceptual Problems

In this section, we will mention several different aspects and features of criticism of Kripke's theory of truth that are connected with conceptual aspects of his theory. The strength of the arguments is sometimes quite dependent on the personal preferences of philosophical background theories, or to put it differently, the strength is dependent on the general philosophical positions of the authors.[16] Examples are whether someone believes in logical pluralism or not, whether someone believes that a priori truths exist or not, or what the preferred theory of meaning (extrinsic, intrinsic, something in between etc.) is. Dependent on these controversial positions some kinds of criticisms question the whole project of Kripke of which some are not very compelling. Keeping this as our background assumption, the considerations in this section are quite relative concerning their strength.

One of the most important features of Kripke's account is the fact that in order to construct the partially defined truth predicates, one has to avoid classical logic. Instead of classical logic, one has to adopt one or the other form of a non-classical many-valued logic with the crucial property to be monotone. In this part of the work, we developed an account as to how to formulate Kripke's ideas in a four-valued logic that can be interpreted as an extension of the three-valued logic based on the strong-Kleene tables.

Many researchers think that using non-classical logics is a counterintuitive move of solving open problems by creating new problems. In their opinion, it is highly counterintuitive to reject obvious tautologies like $p \rightarrow p$, $p \vee \neg p$, or

[16] As most philosophical positions they are most often highly controversial.

$\neg(p \wedge \neg p)$. At the same time, they claim that it is highly counterintuitive to accept cases where $\neg(p \vee \neg p)$ is not a contradiction. Some people (in the newer history [Bo97]) would count the above tautologies as a priori truths (respectively an a priori contradiction like the last example). Despite the fact that it is still highly controversial which statements count as a priori truths, it is clear that the intuitive correctness of the above principles has some power. It is a matter of fact that in ordinary common sense reasoning, most often we reason using these logical laws. Whether or not we believe that these principles are a priori true or only an empirical generalization of experience, using these principles leads us to quite appropriate models of the world. And under the assumption that these principles are successful principles, then a theory that works with a logic that does not validate these (or validates the negation in the case of the expression $p \wedge \neg p$) must be a bad theory. Notice that in the three-valued case as well as in the four-valued case of our construction these principles are not valid. Is this a sufficient reason to reject Kripke-style constructions?

In order to give reasonable criteria to answer this question, it is important to notice that there is a large number of different disciplines in philosophy, information theory, computer science, and cognitive science in which non-classical logics are used and successfully applied in order to model reality.[17] For example, non-classical logics are quite often used to model quantum physics (so-called ortho-logics), knowledge representation (many-valued logics), belief-revisions (non-monotonic logics), common sense reasoning (abductive reasoning), presuppositions (dynamic or many-valued logics), semantics of natural languages (a variety of different types and techniques of non-classical logics), or information transfer (again a number of different techniques and types).[18] It is highly unlikely (some would probably say impossible) that in the future we will be able to give representations of these disciplines without a usage of non-classical logics. Notice that the problems which arise if one tries to use classical logics are not technical ones, but most often problems on a principal level. Notice further, that it is not only empirical experience that tells us that classical logic is not optimal for modeling truth predicates of natural languages, but provable mathematical theorems that prevent us from being successful (for example Tarski's theorem). It is simply impossible to represent the truth predicate for natural languages using a completely classical framework (that means classical logic, classical syntax, classical semantics etc.). That is a strong argument in order to reject the criticism that claims that non-classical frameworks for a theory of truth are not appropriate.

It is clear that the remarks we mentioned cannot be sufficient for a complete discussion of this point. Moreover, there are deeper philosophical problems involved, namely such problems as a priori knowledge, empirical knowledge, relativism, and last but not least a philosophical discussion of

[17] Interestingly enough even in mathematics there was the attempt to give an alternative logical foundation where some of the classical tautologies do no longer hold. This alternative theory is intuitionistic mathematics.

[18] From a more abstract point of view, the main problems of classical logic if applied in these contexts are the bivalence principle and the monotonicity feature. Clearly all the mentioned logics have different properties, but often they try to give solutions for avoiding these features.

what truth really is (implying problems like idealism, realism, externalism, internalism etc.). These problems are highly controversial and cannot be discussed informally. We will not consider the deeper context of this discussion because that cannot be the focus of this work. It is clear that in this work, non-classical accounts for logics, semantics, and set theory (like revision theories, many-valued logics, or non-classical set theory (circular set theory)) will be used quite often throughout the frameworks of circularity. It is in the spirit of this work to assume that there are useful applications of non-classical logics, at least in the realm where those alternatives yield appropriate solutions.

Another problem for Kripke's account concerns criteria for the choice of the correct fixed point in order to get the 'right' extension of the truth predicate. What are sufficient and necessary criteria to choose this predicate? Notice that for a countable language there are infinitely many truth predicates if one uses the strong-Kleene tables of the construction based on a three-valued version (as well as in the case of the four-valued case).[19] We have a lot of possible truth predicates that can be taken as the correct ones. Kripke's original proposal was - although he was not very clear concerning this point - to choose either the minimal fixed point or the maximal intrinsic fixed point of the construction. Even if we have a hint as to which fixed point we can choose, we will see later that that hint is not very reliable. If there is no interesting maximal intrinsic fixed point - as in the construction presented in this work - the problem still remains how to choose a particular extension of the truth predicate from infinitely many possibilities. Although there are clearly some negative hints - for example, that a maximal fixed point would not be a good choice (because the truth-teller would be always true or false in the three-valued case and both, true and false in the four-valued case - it is less clear to argue for a specific extension and to give positive criteria for a particular point. That kind of criticism is generally accepted and makes a conceptual weakness visible on a principal level of Kripke's account. The only alternative is to choose extensions of the truth predicate according to the purpose of the applications. Although this is extremely unsatisfying for the general theory and the philosophical basis, it is a way out in particular applications of Kripke's theory. In the general case, there is not even no idea as to what kind of criteria are appropriate.

Besides these problems arising in applications of the fixed point approach, there is a further concern on a very deep level. Kripke's account cannot tell us what the meaning of truth is. A theory of truth should give an account that fixes the significance of the truth predicate in every possible world.[20] From a metalevel this means that the significance of truth is fixed by Tarski's biconditionals in all possible worlds. It is relatively hard to give good arguments against this thesis, because Tarski's biconditionals provide precisely the intuitive meaning of the truth predicate.

[19] Compare [CaDa91].

[20] Anil Gupta calls this requirement the signification thesis and uses signification in a technical sense. Cf. [GuBe93].

Exactly the signification thesis does not hold in the fixed point approach: if we have only two extensions, i.e. fixed points of the truth predicate (and in Kripke's account there are infinitely many extensions of the truth predicate), then the claim that the signification of the truth predicate is fixed by Tarski's biconditionals is necessarily false as Gupta shows in [GuBe93]. This consequence is highly counterintuitive if one wants to avoid relativism concerning the truth predicate. It is probably not correct to claim that the thesis of relativism concerning the truth predicate is not existent in philosophical disputes. The majority of philosophers would agree that the uniqueness and the determination of the truth predicate by Tarski's biconditionals is one of the very basic requirements of a theory of truth. Precisely this cannot be achieved by Kripke's fixed point approach.

In the next section, we will consider some ontological assumptions concerning Kripke's fixed point approach. We will mention again some remarks concerning many-valued logics. Furthermore, we will consider some problems dealing with the bearer of truth and the inductive process essentially used in fixed point accounts.

6.4 Some Ontological Considerations

In order to model a phenomenon, one has to use certain tools and theories taken from mathematics or other sciences. Usually, it is taken as an intuitive truth (sometimes even as an a priori truth) that the ontological assumptions should be as weak as possible. If two theories can explain a phenomenon equally correctly, then the better theory is the theory that is easier and has weaker ontological assumptions.[21]

An important point concerning theories (frameworks) of circularity is the question which assumptions are necessary for such a theory.[22] In the following subsection, we will consider more closely Kripke's assumption of a non-classical logic, the infinite process in order to model a Kripkean-style truth predicate, and the old question whether sentences or propositions are the bearers of truth.

6.4.1 Three-valued Logic

The special feature of Kripke's account is the usage of Kleene's three-valued logic in order to give an account of a theory of truth. The idea that classical logic is sometimes not sufficient in order to model the world is probably more than 2000 years old. For example, Aristotle argued that there is no chance to give a definite truth value to a sentence which makes a claim about the future. The standard modern solution for this problem is to adopt one or the other form of tense logic.[23] Although most versions of tense logic are relative

[21] There is no consensus how the situation must be evaluated in a case where one theory is easier to handle but the other theory has weaker assumptions. Here, one has to decide from case to case.

[22] We will see later that these assumptions can become quite strong in certain frameworks.

[23] For an overview compare [Be83].

straightforward extensions of ordinary modal logic (that is itself an extension of classical propositional logic), it is nevertheless not a classical logic we are dealing with when we reason about problems like the one Aristotle mentioned. New operators (for example Kamp's S and U predicates[24]) must be introduced. The formal properties of these tense logics (as well as of modal logics) vary. Many of them are complete and decidable (for example S and U tense logic), others are not. Many (propositional) modal logic formulas can be interpreted as formulas of first-order predicate logic, some of them cannot. One has to be aware that the logics we are dealing with are not classical logics. They have different properties and are a proper extension of classical logic.

What commitment do we have when we adopt Kripke's ideas? First, we mention that Kleene's three-valued logic as well as the four-valued approach described here using bilattices is equivalent to classical logic if we restrict the truth values to the set $\{T, F\}$. In other words, we do not change anything when we reason with ordinary truth values.[25] Differences arise only when we deal with sentences that do not have any definite truth value. Therefore, this step does not seem to be problematic at all.

There are features of the described many-valued logics that are usually not wanted: in the three-valued version as well as in the four-valued version, there are no tautologies. What commitment do we have concerning this point? Do we have to presume that there are no tautologies out there in the world? I think that there is no trivial answer to this question. First, one has to be aware that not all proof systems can be adopted for many-valued logics. Because the described three-valued and four-valued logics do not have tautologies, a Hilbert style system cannot be adopted because these calculi have tautologies as axioms. Gentzen type proof systems fit better into the context. As a matter of fact, it is possible to adopt a proof calculus for these logics. Therefore, the concept of proof makes sense even in these non-classical logics.

Second (and as a consequence of the first point), standard principles of classical logic do not longer hold in Kripke's account. For example, the principle of the excluded middle is no longer valid in three- (or four-) valued logic. For some researchers, this consequence is not acceptable, especially for people who claim that such logical principles are in fact a priori truths.[26] Usually, principles like the principle of the excluded middle are used for nearly all kinds of reasoning. Especially, this principle plays an important role for reasoning in the meta-language, for example, in mathematical proofs. We have to agree that non-classical logics can only be used to describe a phenomenon, but that the reasoning on a metalevel about this non-classical logic must be done in ordinary classical logic.[27] This is at least a common basis for skeptics concerning

[24] Cf. [Bu79].

[25] An important point is that the information order is vacuous in the case we restrict our examination to the two classical truth values. Without one of the truth values N or B the framework CCPO, CPO, or bilattice makes no sense.

[26] It is an implausible simplification that these people claim something to be an a priori truth without having an appropriate theory of truth. It is simply begging the question of the real problem. But it is not in the focus of this work to work out a criticism of theories claiming the existence of a priori truths.

[27] Intuitionistic logic is the only developed account of a logic that can be used on an object-

non-classical logics and for logical pluralism. Clearly, this does not prevent us to take a position for or against the usage for non-classical logics.

Third, non-classical logics are usually only developed on a propositional level. Non-classical versions of predicate logic often do not exist. Examples are the low level of knowledge concerning quantified modal logic, quantified tense logic, and quantified many-valued logics.[28] An exception is again intuitionistic logic. Here, it is relatively well understood what happens if one introduces quantification over individual variables. It does not seem to be appropriate to use a logic without knowing the properties of the logic when one introduces quantification over variables. From an ontological perspective, a possible reason for the emphasis on the propositional non-classical logics is the relative strength of these logics. Quite often, they are completely sufficient to model the objects we want to model. Then, there is no real motivation to go further.

Clearly, this kind of criticism is not a very strong argument against Kripke's account. Dependent on the background assumptions, Kripke's construction is problematic because of the ontological assumption of a many-valued logic. For logical pluralists, the pure possibility of modeling the truth predicate of natural language by Kripke's account is an argument for the usage of non-classical logics.

6.4.2 The Inductive Process

A further point concerns the inductive process in the construction of the truth predicate. We have to assume that there is a truth predicate for natural language which is defined in an infinite process. Obviously, human capacities are strong enough to deal with a truth predicate. On the other hand, there is no chance to get the correct extension of such a predicate if this extension can only be achieved via an infinite process. We are finite and every attempt to perform the definitional process must necessarily fail. Humans do not have the cognitive capacity to perform an infinite inductive process.

We have to distinguish two levels. On the one hand there is the understanding and the usage of the truth predicate for natural language and on the other hand there is the modeling of this truth predicate in a formal framework, using mathematical tools. The usage of the truth predicate is determined through situations in the world, with the strong tendency to generalize everything that seems to be regular. We are working in the idealized world of mathematics in order to get a precise concept of the extension of the truth predicate. Both dimensions must be clearly distinguished and need to be kept separate. The strategy to define the truth predicate in an infinite inductive process does not mean that the ordinary user of this truth predicate is in fact successfully performing this inductive process in order to get the ability to use the concept

and metalevel. The special form of intuitionistic mathematics has different properties in comparison with ordinary mathematics. Compare [TrDa88], [KlVe65], or [He71] for an overview of the differences between intuitionistic mathematics and ordinary mathematics.

[28] Clearly the term 'low level' must be interpreted relative to the knowledge of predicate logic.

truth predicate in a correct way. In other words, modeling is not the same thing we are performing when we use the truth predicate in our reasoning.

From this perspective, it does not seem to be the case that ontological assumptions are too strong for a correct analysis of the truth predicate. To use a monotone four-valued logic in order to model circularity is simply a flexible possibility to get an approach towards a theory of truth, not a theory that tells us what humans do when using the truth concept for their reasoning.[29] If one keeps these levels apart, then no problem should arise from Kripke's account concerning this point. The infinite inductive process is simply a possible representation of the truth predicate, not the representation of the processes in the brain when humans use the truth predicate.

6.4.3 Sentences vs. Propositions

A further problem of Kripke's account concerns the bearer of truth. Whereas historically many attempts to model pathological sentences assume that sentences are the bearer of truth, this is less clear since the modeling of natural language has become more and more successful in giving a semantics for natural language processing. One important aspect of linguistics takes into account the contextual dimension of natural language, including indexicals, and the situation in which a sentence is uttered.[30] That these contextual features are important concerning the treatment of pathological sentences is also quite obvious. Not only a single sentence can be pathological but a whole discourse can be pathological as well. Necessarily, one has to take into account in what situation a single sentence is uttered if one tries to give a treatment of pathological discourses. Additionally, as newer treatments of pathological sentences show, a fine structure of the very same sentence is an essential feature of the analysis (compare Chapter 9 of this work). Finally, as even Kripke's own examples in [Kr75] show (and the examples in the previous chapters of this work), we have to take into account who is uttering a pathological sentence, when such a sentence is uttered, who the audience is, and what background knowledge is given in a certain discourse. In total, the context is an important ingredient in the analysis. This claim cannot be reasonably disputed.

Given this feature of truth theories, it is less clear whether a treatment that is completely focused on sentences as the bearer of truth can give a good analysis of pathological sentences on a principal level. It is quite probable that, in order to give an appropriate treatment of such sentences, one has to give an account that assumes that propositions are the bearer of truth. This means that not only the sentence itself must be coded in the modeling but additionally,

[29] In many philosophical discussions, it seems to be the case that there is a confusion between the cognitive capacity on the one hand and the modeling of this capacity on the other. An example is the early history of transformational grammar, where many linguistics thought, that humans really have trees in their mind, when they produce a sentence. Today nobody would claim this. Trees are simply a possibility to model regularities of language, not to describe what we do, when we use language.

[30] In a certain sense, the complete development of the semantics of natural language in the last twenty years deals with this problem. Examples of theories for contextual properties of natural language are [KaRe93, ViVe96, BarPe83, BarEt87], or [Hei82].

for example, the time when the sentence is uttered, the person, who utters the sentence, and the context in which the sentence is uttered. We do not claim that the question whether to use a theory of propositions or a theory of sentences is purely a question concerning contextual aspects. These problems are rather independent from each other. But we try to make clear that it is important in which world or in which situation a particular sentence is uttered.

To develop a theory of propositions is much more complicated than a theory that restricts its scope to sentences. The reason for this is because sentences can be regarded as inductively defined from elementary relations and the ontological status of such entities is quite uncontroversial as well - in contrary to the ontological status of propositions).[31] We conclude that for an understanding of the human truth concept in many examples it is necessary to take propositions into account. We do not claim that this is essential for every application, but we claim that a complete theory of truth for natural language must include a theory of propositions.

We finish this section of ontological criticism of Kripke's account with these remarks. The discussion was focused on some important points. For further criticism the reader is referred to the literature in the history section.

6.5 History

As well as the number of accounts to model truth predicates in general, even the literature that just criticizes Kripke's approach is enormous. A good overview of the discussion in analytic philosophy concerning Kripke's fixed point approach can be found in [Ma84]. In this work, different articles are collected that discuss and criticize the fixed point approach on a philosophical level as well as on a technical level. Other sources of material for criticism of partially defined truth predicates are [GuBe93, BarEt87, Mc91, Bre92], and [Ki92]. An important influence was the criticism of Gupta (in [Gu82]). This work, together with ideas of Herzberger, initiated a further branch of the development of truth theories, namely so-called revision theories, originally developed in [Gu82, He82a] and [He82b] (compare Part III of this work). We restricted the discussion to relatively abstract topics concerning problems of Kripke's approach. Another aspect of the criticism of Kripke's theory concerns his claim that he gave an account for a truth predicate of natural language. It is not really clear what properties such a truth predicate must precisely have, but it is quite clear that Kripke's theory cannot be a candidate for such a modeling (at least in the form he proposed). Probably that is the only agreement amongst the numerous commentators of partially defined fixed points.

[31] At this point we are not able to give a sufficient theory of propositions. Later in this work we will come back to this point and we will discuss this more precisely. Compare Section 15.2.

Part III

Generalized Definitions: Gupta-Belnap Systems and their Applications

Chapter 7

Definitions and Games

In the last chapters, we modeled self-referential sentences in a non-classical four-valued logic. This account has the advantage that the underlying set theory, the definition theory, and the semantics (the model theory) of the framework remain classical. The only parameter that is changed in order to get a framework for circularity is the underlying logic. Another possibility to model circularity is to work in classical logic, but to change one (or more) of the other features of the formal framework, namely the semantics, the definition theory, or the underlying set theory. In this third part, we will examine generalized definition theories, i.e. definition theories that allow us to define predicates in a circular way. In other words, we allow that the definiendum can occur in the definiens. The following formula (in which H is a given predicate and D denotes a given domain) is an example of a circular definition:

$$\forall x \in D : (G(x) \ \Leftrightarrow \ \neg G(x) \vee H(x))$$

Notice that in classical definition theory such a definition is ill-formed: in a classical definition it is not possible that the definiendum is included in the definiens. Notice further that in classical definition theory, it is not possible (in general) to assign an extension to G. For example, if $H(x)$ does not hold for any $x \in D$, then it holds for all $x \in D : G(x) \ \Leftrightarrow \ \neg G(x)$. Clearly, no extension can be assigned to G in classical definition theory. Intuitively, we should be able to assign an extension in the case $H(x)$ does hold for every $x \in D$. In this case, $G(x)$ should (intuitively) hold for all $x \in D$, too.

In order to avoid the problem of becoming inconsistent, it is not only necessary to allow circular definitions but furthermore to develop a model theory for predicates that are defined using circular definitions. The semantics of the framework must be changed in one important respect in order to be able to get an appropriate interpretation of arbitrary circular definitions: we need to find a way to evaluate circular definitions consistently. The development of such a theory is the heart of the so-called Gupta-Belnap systems we will consider in this part of the work. The attempt tries to show that it is possible to find a semantics for circular definitions.

In this chapter, we will begin our considerations with first-order definitions and the corresponding game theoretical characterizations. Additionally, we will introduce inductive definitions and we will give a characterization in game theoretical terms. This chapter can be interpreted as preliminaries for revision theories we will consider in Chapter 8.

7.1 Preliminaries

First-order theories are well understood and sufficiently enough developed. The general question concerning first-order theories is to give criteria in order to characterize classes of structures that are first-order definable, i.e. that can be defined using a first-order condition ϕ. For example, the property of being a structure that is non-empty and that contains at least two elements can be represented as a first-order definition. To make this precise, assume $\mathcal{A}$ is a given structure. Then, the following definition provides a characterization of the property of structure $\mathcal{A}$ not being empty with cardinality bigger than 1.

$$G(\mathcal{A}) \;\Leftrightarrow\; \exists x_1 \in |\mathcal{A}| \exists x_2 \in |\mathcal{A}| : x_1 \neq x_2$$

The above definition of predicate G does precisely what we want: the domain of $\mathcal{A}$ has at least two distinct elements. Unfortunately, the expressive power of first-order definitions is rather limited. If one tries to define the natural numbers, one recognizes that this cannot be done by a first-order formula. The same is true for the collection of all connected finite graphs, the collection of all finite palindromes (relative to a given finite alphabet), or the collection of all arithmetically true sentences. All these examples require stronger techniques and methods in order to be precisely defined, because they are higher-order concepts. Therefore, if one tries to strengthen formal systems (theories), it is a good strategy to extend definition theories in one or the other respect. We give some examples how this can be achieved.

One possibility to strengthen a theory is to define new predicates (functions or relations) using inductive definitions, i.e. definitions that are well-founded on a base case (or several base cases), such that an iterative process yields an extension of the predicate. In this context, the iterative process means that every stage $n+1$ can be calculated using the result from stage n. The algebraic counterpart of an inductive definition is the minimal fixed point of a monotone operator. The fixed point gives the interpretation of the newly defined predicate.[1] In other words: given a certain monotone operator Γ, it is possible to identify the minimal fixed point of this operator with the extension of an inductively defined predicate. We can give an easy example of this concept.

Example 7.1.1 We want to define the natural numbers $\mathbb{N}$. The following formula does precisely this.

[1] Non-monotone inductive definitions interpreted as the minimal fixed point of a non-monotone operator is another possible extension, but knowledge about this kind of strengthening a language is rather limited. It should be noticed at this point that generalized definition theory as will be developed in this chapter is not a theory of non-monotone inductive definitions.

$$n \in \mathbb{N} \Leftrightarrow \forall S(\forall x(0 \in S \land (x \in S \to x' \in S)) \to n \in S)$$

Notice that we only assume that a certain successor function $'$ and the base case element 0 are given. We do not assume that the successor function is defined as the application of the power set operation. We can choose other codings for the successor function as well.

Kripke's construction is essentially a fixed point construction using a monotone operator in order to get fixed points that determine possible extensions of the truth predicate. We saw that the choice of an appropriate fixed point in Kripke's framework is a problematic issue. The minimal fixed point corresponds to an inductive definition of the truth predicate.[2] Besides the minimal fixed point the maximal intrinsic fixed point plays a very important role in Kripke's theory of truth. Whereas the minimal fixed point can count as a straightforward example of an inductive definition this is not trivially the case for the maximal intrinsic fixed point.

Because of the fact that monotone operators have (in general) many different fixed points, it is a natural idea to call a predicate P coinductively defined if P is defined via the maximal fixed point of a given operator. Algebraically, the idea is to use the dual of the inductive definition in order to get a new construction method. We will consider coinductive definition (as well as inductive definitions) later in Section 7.3 and Section 7.4.

Although all these extensions strengthen the power to define a predicate in a language L, there is no possibility to define a fully circular concept like the Liar sentence λ in L. The reason for this is the fact that circular definitions are ill-formed expressions according to the classical understanding what a definition really is. Technically, a definition is well-defined and determines the extension of a predicate P (in the classical theory) if it holds for all models $\mathfrak{M} = \langle D, I \rangle$ of a given language L:

$$\exists \phi \in L \forall x \in D : (x \in P \Leftrightarrow \models^{\mathfrak{M}} \phi(x))$$

The formula ϕ must not contain the predicate P. But ϕ can contain other predicates for which an extension (an interpretation) is already fixed. If ϕ is a first-order formula, it cannot contain second-order quantifiers over predicates and relations. Quantifiers can only range over individual variables. Dependent on the complexity of the quantification, the definition can be classified in complexity classes according to the Kleene (or arithmetical) hierarchy.[3]

Let us consider a circular definition for a moment. An example of a circular definition is the formula from above.

$$\forall x \in D : (G(x) \Leftrightarrow \neg G(x) \lor H(x))$$

[2] As the base case one takes the bottom $\perp$ whereas the inductive process is generated by the monotone operator Γ. As can easily be proven one reaches finally a fixed point.

[3] Compare Chapter 1 for further information.

This definition contains the predicate G (that shall be defined) on the right side of the definition. It is therefore ill-formed in classical terms. The reason to restrict a definition in the classical theory is to prevent the case where it is impossible to assign extensions to some circularly defined predicates. Notice that the above definition does not have an appropriate (classical) extension if $H(x)$ does not hold for any $x \in D$.

The impossibility to assign an extension for G remains also true if one tries to extend the theory using inductive (or coinductive) definitions. It is immediately clear that there is no monotone (even no non-monotone) operator available defining an appropriate fixed point for the extension of a predicate that has the properties of the Liar sentence. More precisely, there is no fixed point at all for such Liar-like definitions. One recognizes immediately the limits of classical extensions of first-order definability. Modeling circular phenomena using circular definitions is a way out of these restrictions. The so-called Gupta-Belnap-Herzberger revision theories (we call these theories in the following either Gupta-Belnap systems (GB-systems) or simply revision theories) are a way out of this dilemma. Gupta-Belnap systems allow an extension of classical model theory giving an appropriate interpretation and semantics (model theory) for circular definitions.[4] Additionally, Gupta-Belnap systems behave for classical definitions like ordinary model theory. That is the reason why Gupta-Belnap systems are an extension of the classical theory.

In this chapter, we will consider certain properties of first-order theories as well as certain different kinds of extensions of first-order theories. We will introduce these extensions step by step. The characterization method of first-order theories is given by the so-called Ehrenfeucht-Fraïssé games where an association of first-order formulas on the one hand and winning strategies in a finite game of two players on the other is possible. Second, we shall extend structures using inductive and coinductive definitions. The game theoretical counterparts for these definitions are open and closed games. As a side remark we mention that the techniques used in this chapter are quite different in comparison with the techniques used in Part II. To begin with we examine the limits of the non-second-order case, namely first-order definitions and their definitional strength using game theoretical techniques and characterizations.

7.2 First-Order Definability

First-order logic deals with relations and predicates that are definable by quantifications over individual variables of a given domain. A quantification over higher-order objects (for example relations, functions, or relations of relations etc.) is not allowed. First-order logic can count as a natural way to extend classical propositional logic. Another possibility for an extension would be the

[4] This does not mean that paradoxical sentences are true in some of these theories. According to the characterization below, we will find that Gupta-Belnap's semantical system $\mathbf{S}^*$ has complexity Π^1_2 and can therefore be defined in a well-founded second-order logic, too. In other words: we do not loose classical mathematical properties.

introduction of further operators like modal, epistemic, or tense operators. Or one introduces intensional operators in order to strengthen first-order logic. Our considerations will be focussed on first-order predicate logic. We begin our examination with some properties and classes of structures that can be defined in first-order logic.

Example 7.2.1 The following properties and classes can be defined in first-order logic.

- Every finite structure can be characterized in first-order logic (up to isomorphisms).
- The property of a structure to be a strict linear order is first-order definable.
- The property of a structure to be non-empty is first-order definable.
- The class of all graphs that are totally connected is first-order definable.
- The class of all graphs that are totally unconnected is first-order definable.
- The property of a given function to be monotone is first-order definable.
- ZFC set theory is first-order definable (using $\in$ as unique non-logical symbol and taking sets as individual variables).

It is obvious in most cases that the claims of Example 7.2.1 do hold. The only claim that requires a bit more reasoning than writing down the defining first-order formula is the first example. For a proof of this example the reader is referred to [EbFl95].

An important feature of first-order logic is the fact that it is a complete logic, and that the compactness theorem, Robinson's theorem, as well as the Löwenheim-Skolem theorems, and the Lindström theorem hold. These features are not necessarily preserved in extensions of first-order logic. For example, in second-order logic, it is well-known that the compactness theorem does no longer hold. Perhaps that is a further reason why first-order logic is the most prominent logic in the history of the foundations of mathematics.

In first-order logic, it is not possible to quantify over functions, relations, relations of relations, or functions of functions. In other words, it is not possible to quantify over higher-order objects. Therefore, our expressive power to define predicates is rather limited. In order to introduce the game theoretical Ehrenfeucht-Fraïssé method, we will work with a running example in this section. We will show that the set of all finite palindromes of a given finite alphabet (in our case the alphabet $\{a, b\}$) is not first-order definable. The example in question can be precisely formulated as follows.

Example 7.2.2 Assume the alphabet $\{a, b\}$ is given. We want to define the set of all finite palindromes *Pal* (i.e. the set of all words, such that each word when read from the left to the right is equivalent to the word when read from the right to the left). As we will see in this section, it turns out that Pal is not definable in first-order logic. That means that there is no first-order formula ϕ, such that it holds:

$$\forall x : x \in Pal \;\Leftrightarrow\; \phi(x)$$

The reason for the fact that the collection of all finite palindromes of a given finite alphabet is not first-order definable is a consequence of the limited expressive power of first-order definitions. We will see in this section that the above fact can be proven using Ehrenfeucht-Fraïssé games. Additionally, we will mention several other examples of non-first-order definable sets. Intuitively, the problem is based on the fact that we want to define the smallest class (the minimal fixed point), such that the elements of this class are precisely all finite palindromes. In other words, we need to define finiteness of palindromes and that turns out to be quite problematic.

To justify the claim of Example 7.2.2 we need some definitions, concepts, and facts. The crucial step for a characterization of first-order formulas (and as a side effect an effective method to check whether a certain property of structures or the definition of a class of structures is first-order definable or not) will be the introduction of Ehrenfeucht-Fraïssé games that allow an easy and intuitive characterization of the concept *similarity of two structures up to a certain degree.*[5]

We begin with a technical definition of the concept of a partial isomorphism. This concept is important to define winning strategies in Ehrenfeucht-Fraïssé games.

Definition 7.2.1 *Assume two structures $\mathcal{A}$ and $\mathcal{B}$ are given (relative to a given language L). We call a function f a partial isomorphism if $dom(f) \subseteq |\mathcal{A}|$, $ran(f) \subseteq |\mathcal{B}|$, and additionally the following conditions hold:*

(a) f is injective.
(b) For all constants $c \in Const$ it holds: if $c \in dom(f)$ then $f(I_{\mathcal{A}}(c)) = I_{\mathcal{B}}(c)$ (where $I_{\mathcal{A}}$ and $I_{\mathcal{B}}$ are interpretation functions)
(c) For each n-ary relation R and arbitrary $a_1, \ldots a_n \in dom(f)$ it holds:

$$||R(a_1, a_2, ..., a_n)||^{\mathcal{A}} = T \iff ||R(f(a_1), f(a_2), ..., f(a_n))||^{\mathcal{B}} = T$$

As we will see below, partial isomorphisms do not necessarily preserve arbitrary quantified formulas. They clearly preserve atomic formulas as well as propositional formulas. Partial isomorphisms f are isomorphisms that are

[5] Ehrenfeucht-Fraïssé games are based on the work [Fr54] and [Eh61]. Whereas Fraïssé introduced the concept of m-equivalence, it was Ehrenfeucht who invented the game theoretical characterization. A good source for further information concerning Ehrenfeucht-Fraïssé games can be found in [EbFl95].

restricted to the subset of the domain of f where they are defined and to a certain subset of the range of f. Partial isomorphisms are natural generalizations of ordinary isomorphisms, namely isomorphisms that are locally defined.

The following definition describes the basic form of an Ehrenfeucht-Fraïssé game between two players I and II.

Definition 7.2.2 *(i) Assume two structures $\mathcal{A}$ and $\mathcal{B}$ are given, where $A = |\mathcal{A}|$ and $B = |\mathcal{B}|$. An n-round Ehrenfeucht-Fraïssé game $G_n(\mathcal{A}, \vec{d}, \mathcal{B}, \vec{e})$ for $\vec{d} = \langle d_1, d_2, \dots, d_n \rangle$ and $\vec{e} = \langle e_1, e_2, \dots, e_n \rangle$ between player I and player II is a sequence of moves of the two players that satisfies the following rules:*

- *Player I starts the game by choosing one of the structures $\mathcal{A}$ or $\mathcal{B}$ and playing a point of the chosen structure $d_1 \in A$ or $e_1 \in B$.*
- *Player II replies by playing a point in the alternative structure: $e_1 \in B$ in case player I played $d_1 \in A$ and $d_1 \in A$ in case player I played $e_1 \in B$.*
- *In the m^{th} round, player I chooses one of the structures $\mathcal{A}$ or $\mathcal{B}$ and plays a point $d_m \in A$ or $e_m \in B$.*
- *Player II replies by playing a point in the alternative structure, such that player II plays $e_m \in B$ in the case player I played $d_m \in A$ and player II plays $d_m \in A$ in the case player I played $e_m \in B$.*

(ii) We say that player II wins the n-round Ehrenfeucht-Fraïssé game $G_n(\mathcal{A}, \vec{d}, \mathcal{B}, \vec{e})$, if there is a partial isomorphism

$$f : \{d_1, d_2, ..., d_n\} \longrightarrow \{e_1, e_2, ..., e_n\}$$

Elsewhere player I wins the game.

(iii) Player II has a winning strategy in the n-round Ehrenfeucht-Fraïssé game $G_n(\mathcal{A}, \vec{d}, \mathcal{B}, \vec{e})$, if it is possible for player II to win each game, whatever choices are made by player I during the game.

Remark 7.2.3 Quite often player I is called the spoiler and player II is called the duplicator. We do not use these notions here, because the references are clear when we call the two players player I and player II.

We will use Ehrenfeucht-Fraïssé games to show that the collection of finite palindromes of the finite alphabet $\{a, b\}$ is not first-order definable. Because of this running example, we do not add many examples for Ehrenfeucht-Fraïssé games. Nevertheless, some examples should be mentioned to get a better understanding of the concept.

Example 7.2.4 (i) Assume $\mathcal{A} = \mathbb{N}$ and $\mathcal{B} = \mathbb{N}$. For every Ehrenfeucht-Fraïssé game $G_n(\mathcal{A}, \vec{d}, \mathcal{B}, \vec{e})$ player II has a winning strategy: if player I plays in the i^{th}

round the element d_i, then player II responds by playing e_i, such that $d_i = e_i$. Clearly, the function f, such that

$$f : \{d_1, d_2, \ldots, d_n\} \longrightarrow \{e_1, e_2, \ldots, e_n\} : d_i \longmapsto e_i$$

is a partial isomorphism between $\mathcal{A}$ and $\mathcal{B}$.

(ii) Assume $\mathcal{A} = \langle \omega, \leq \rangle$ and $\mathcal{B} = \langle \omega + \omega, \leq \rangle$ are given (where $\leq$ is the standard order relation on the ordinals and $\omega + \omega$ is interpreted as order theoretic sum). For every fixed $n \in \mathbb{N}$ player II has a winning strategy in the game $G_n(\mathcal{A}, \vec{d}, \mathcal{B}, \vec{e})$. In order to see this, one has to check several cases. We mention only the general idea. The problematic case is when player I plays in the j^{th}-round an element $e_j = \omega + k$ in the structure $\langle \omega + \omega, \leq \rangle$. Player II has to respond by playing an element d_j in structure $\langle \omega, \leq \rangle$ that is big enough in order to respond to possible moves of player I in the rounds $j + 1$ to n. This can be achieved by playing an element $d_j = x + 2^{n-j}$ where x is dependent on the moves performed in round 1 to round $j - 1$. Notice that n has to be fixed in advance. It does not hold that player II has a winning strategy for a game in which $n \in \mathbb{N}$ is not fixed.

(iii) If $\mathcal{A}$ is a totally connected graph and $\mathcal{B}$ is a graph that contains an isolated point e, then player II cannot have a winning strategy in $G_n(\mathcal{A}, \vec{d}, \mathcal{B}, \vec{e})$ for $n \geq 2$. Player I simply has to play the isolated point e (and a further arbitrary point $e' \in B$) such that player II cannot respond (because $\mathcal{A}$ is totally connected).

The mentioned examples give a rough idea of the basic concepts used in Ehrenfeucht-Fraïssé games. Games and the connected winning strategies are a natural description to show that certain structures are structurally equal up to a certain degree. In order to be able to formulate this fact, we mention a well-known result that connects formulas that hold in given structures $\mathcal{A}$ and $\mathcal{B}$ (relative to a given language L) and n-round Ehrenfeucht-Fraïssé games, respectively winning strategies of player II. This result will be formulated (and proven) in Theorem 7.2.6. In order to be able to show this result, we need to introduce the concept of the depth of a formula, and the idea of a preloaded game. The following definitions specify these two concepts precisely.

Definition 7.2.3 *The depth of a formula ϕ is inductively defined as follows:*[6]

depth$(\phi) = 1$*, if* ϕ *is an atomic formula*
depth$(\phi \wedge \psi)$ = *max(depth*(ϕ)*,depth*(ψ)*)*
depth$(\neg\phi)$ = *depth*(ϕ)
depth$((\forall x)\phi)$ = *depth*$(\phi) + 1$

[6] Notice that we can define other logical connectives by the semantics of $\neg$ and $\wedge$. Note further that an existential quantified formula is definable via the universally quantified formula according to the well-known quantifier rules of predicate logic.

Definition 7.2.4 *We call the n-round Ehrenfeucht-Fraïssé game $G_n(\mathcal{A}, \vec{d}, \mathcal{B}, \vec{e})$ preloaded with $\langle d_1, d_2, \ldots, d_k\rangle$ and $\langle e_1, e_2, \ldots, e_k\rangle$, if $d_1, d_2, \ldots, d_k \in A$, and $e_1, e_2, \ldots, e_k \in B$ and additionally it holds:*

$$\begin{aligned} &G_n(\mathcal{A}, \vec{d}, \mathcal{B}, \vec{e}) \\ &\qquad = G_n(\mathcal{A}, \langle d_1, d_2, \ldots, d_k, d_{k+1}, \ldots, d_n\rangle, \mathcal{B}, \langle e_1, e_2, \ldots, e_k, e_{k+1}, \ldots, e_n\rangle) \end{aligned}$$

Player II has a winning strategy in the preloaded Ehrenfeucht-Fraïssé game $G_n(\mathcal{A}, \vec{d}, \mathcal{B}, \vec{e})$, if $f : d_i \mapsto e_i$ is a partial isomorphism for $i \in \{1, 2, \ldots, n\}$. Elsewhere, player I has a winning strategy. (The other definitions remain equal to the ordinary definition of an Ehrenfeucht-Fraïssé game.)

We come back to our running example. We need to introduce the idea of n-equivalence. The following definition makes this precise for the special case of words over the alphabet $\{a, b\}$.

Definition 7.2.5 *We define an equivalence relation $\equiv_n$ on letters of finite words over the alphabet $\{a, b\}$ as follows:*[7]

$$\begin{aligned} x_1x_2\ldots x_k \equiv_n y_1y_2\ldots y_k \iff &\forall i, j \in \{1, 2, \ldots, k\} : \\ &(d(x_i, x_j) \leq n \to d(x_i, x_j) = d(y_i, y_j)) \\ &\wedge (d(y_i, y_j) \leq n \to d(x_i, x_j) = d(y_i, y_j)) \end{aligned}$$

Notice, that $x_1x_2\ldots x_k \equiv_1 y_1y_2\ldots y_k$ means that there is partial isomorphism $f : x_i \mapsto y_i$ for $i \in \{1, 2, \ldots, k\}$. Intuitively, the above definition requires that letters that are close enough to each other (namely $\leq n$) need to have the same distance in both structures.

What is needed in order to establish the undefinability of the collection of finite palindromes is a connection between games and structural similarity of two structures formulated via formulas of a certain depth. The following Theorem points out this crucial connection between Ehrenfeucht-Fraïssé games and the structural similarity between two given structures. It is no exaggeration to claim that 7.2.6 is the heart of finite model theory.

Theorem 7.2.6 *Let $\mathcal{A}$ and $\mathcal{B}$ be two structures of the same language L and let $\phi(x_1, x_2, \ldots, x_k)$ be a formula of L of depth m. Then it holds: if player II has a winning strategy in every n-round Ehrenfeucht-Fraïssé game for $n \leq m$, then for all $d_1 \in A, d_2 \in A, \ldots, d_k \in A$, and for all $d'_1 \in B, d'_2 \in B, \ldots, d'_k \in B$ the following equivalence holds:*

$$\mathcal{A} \models \phi(d_1, \ldots, d_k) \iff \mathcal{B} \models \phi(d'_1, \ldots, d'_k)$$

[7] The expression $d(x, y)$ denotes the distance between to letters in the word. It is defined in the usual way. For example, if two letters are neighbors of each other, then the distance $d(x, y)$ of letters x and y has value 1. An inductive process enables us to use all finite distances between letters.

Proof: The proof is an induction on m. The case $m = 1$ reduces to the case where ϕ is an atomic formula. Now, assume that $\mathcal{A} \models \phi(d_1)$. We have to show that $\mathcal{B} \models \phi(d'_1)$. Because player II has a winning strategy in the 1-round game, player II can respond to every move of player I. Therefore, there is a partial isomorphism $f : d_1 \mapsto d'_1$. The definition of a partial isomorphism guarantees that the relations and functions are preserved under this mapping. Therefore, we have $\mathcal{B} \models \phi(d'_1)$ as well. The right to left direction is the dual statement and a simple application of the definition of partial isomorphism.

Assume that for formulas $\psi(d_1, \ldots, d_k)$ of depth $m \in \mathbb{N}$ it holds:

$$\mathcal{A} \models \psi(d_1, \ldots, d_k) \iff \mathcal{B} \models \psi(d'_1, \ldots, d'_k)$$

Now assume that $\phi(d_1, d_2, \ldots, d_l)$ is of depth $m+1$, player II has a winning strategy in every n-round game with $n \leq m+1$ and it holds: $\mathcal{A} \models \phi(d_1, d_2, \ldots, d_l)$. Without loss of generality we can assume that ϕ is of the form $\forall y \psi$. So we have $\mathcal{A} \models \forall y \psi(d_1, d_2, \ldots, d_l)$. It remains to show that it also holds: $\mathcal{B} \models \forall y \psi(d'_1, d'_2, \ldots, d'_l)$. Consider a $m+1$-round game preloaded with $d_1, d_2, \ldots, d_l$ and $d'_1, d'_2, \ldots, d'_l$, and assume player I plays $d'_{l+1} \in B$. Because player II has a winning strategy, player II can respond by playing $d_{l+1} \in A$, such that $d_1 d_2 \ldots d_{l+1} \equiv_1 d'_1 d'_2 \ldots d'_{l+1}$. Therefore, $f : d_i \mapsto d'_i$ (for $i \in \{1, 2, \ldots, l+1\}$) is a partial isomorphism and preserves the structural properties. Together with the induction hypothesis we find: $\mathcal{B} \models \forall y \psi(d'_1, \ldots, d'_l)$. The other direction is proven in a similar way (dual reasoning). q.e.d.

Theorem 7.2.6 provides a characterization of those structures that make the same formulas true up to a certain complexity of these formulas. The possibility to give an analysis of the fine structure is the crucial point in the theorem.

With these preliminaries we consider again Example 7.2.2. We want to show that the set of all (finite) palindromes is not first-order definable (relative to the given alphabet $\{a, b\}$). Consider the following two structures: $a^{2^n} b^{2^n} a^{2^n}$ and $a^{2^n} b^{2^n} a^{2^{n+1}}$. The structures (words) consist of letters and a binary relation R holding between neighbor-letters of the structures and between letters and the empty set. More precisely, $R(x, y)$ holds if x is the left neighbor of y (or similarly y is the right neighbor of x) where $x, y \in \{\epsilon, a, b\}$. Notice that R can represent that x is the left most letter and y is the right most letter.

First, we show that in these two structures player II has a winning strategy in the case the game $G_m(\mathcal{A}, \vec{d}, \mathcal{B}, \vec{e})$ has at most length $m \leq n$. We state the fact as follows.

Fact 7.2.7 *Assume the structures $a^{2^n} b^{2^n} a^{2^n}$ and $a^{2^n} b^{2^n} a^{2^{n+1}}$ are given. Player II has a winning strategy in the m-round Ehrenfeucht-Fraïssé game $G_m(\mathcal{A}, \vec{d}, \mathcal{B}, \vec{e})$ if $m \leq n$.*

Proof: It is sufficient to show that the following fact holds. If $x_1 x_2 \ldots x_k \equiv_{2^n} y_1 y_2 \ldots. y_k$ (for $k < n$), then the following two conditions hold:

- $(\forall x^* \in a^{2^n}b^{2^n}a^{2^n})(\exists y^* \in a^{2^n}b^{2^n}a^{2^{n+1}}) : x_1x_2\ldots x_kx^* \equiv_{2^{n-1}} y_1y_2\ldots y_ky^*$
- $(\forall y^* \in a^{2^n}b^{2^n}a^{2^{n+1}})(\exists x^* \in a^{2^n}b^{2^n}a^{2^n}) : x_1x_2\ldots x_kx^* \equiv_{2^{n-1}} y_1y_2\ldots y_ky^*$.

We will only consider the case where player I chooses x^* located between x_i and x_j. If $d(x_i,x_j) \leq 2n$, then (according to the definition of $\equiv_m$) $d(y_i,y_j) \leq 2n$. Hence, there is a letter y^*, such that $x_1x_2...x_kx^* \equiv_{2^{n-1}} y_1y_2...y_ky^*$. But if $d(x_i,x_j) > 2n$, then $d(y_i,y_j) > 2n$, too. Therefore, we find a letter y^*, such that $x_1x_2...x_kx^* \equiv_{2^{n-1}} y_1y_2...y_ky^*$ as well. There are some other cases to check. This is not difficult to do and we skip it here. The second condition can be shown similarly.

Now, we see that player II has a winning strategy in the n-round game. After the first round we have: $x_1 \equiv_{2^n} y_1$. In the second round, we have the situation $x_1x_2 \equiv_{2^{n-1}} y_1y_2$. After n rounds it holds: $x_1x_2...x_n \equiv_1 y_1y_2...y_n$. In other words: there is a partial isomorphism between the structures $a^{2^n}b^{2^n}a^{2^n}$ and $a^{2^n}b^{2^n}a^{2^{n+1}}$. This completes the proof that player II has a winning strategy in the m-round game for $m \leq n$. q.e.d.

Fact 7.2.7 shows that the representation of differences between the structures $a^{2^n}b^{2^n}a^{2^n}$ and $a^{2^n}b^{2^n}a^{2^{n+1}}$ requires complex formulas with depth $> n$. Fact 7.2.7 is the crucial point in order to show the main claim in this section, namely the undefinability of the set of finite palindromes over the alphabet $\{a,b\}$. We can state the undefinability of finite palindromes of a finite alphabet as a proposition.

Proposition 7.2.8 *The set of finite palindromes over the alphabet $\{a,b\}$ is not first-order definable.*

Proof: To show that the palindromes are not first-order definable assume that the set of palindromes was first-order definable, i.e. assume there is a formula ϕ such that a structure (word) w satisfies ϕ iff w is a palindrome. Then, the defining formula ϕ has a certain depth, say ϕ is of depth n (for $n \in \omega$). Consider again the structures $a^{2^n}b^{2^n}a^{2^n}$ and $a^{2^n}b^{2^n}a^{2^{n+1}}$. Because of Fact 7.2.7, player II has a winning strategy in every m-round game for $m \leq n$. In particular it holds: player II has a winning strategy in the n-round Ehrenfeucht-Fraïssé game. Using Theorem 7.2.6 we get the equivalence:

$$a^{2^n}b^{2^n}a^{2^n} \models \phi \quad\Leftrightarrow\quad a^{2^n}b^{2^n}a^{2^{n+1}} \models \phi$$

Now we have on the other hand, $a^{2^n}b^{2^n}a^{2^{n+1}} \not\models \phi$, because $a^{2^n}b^{2^n}a^{2^{n+1}}$ is not a palindrome and also it holds $a^{2^n}b^{2^n}a^{2^n} \models \phi$, because $a^{2^n}b^{2^n}a^{2^n}$ is a palindrome. We have deduced a contradiction. Therefore, our assumption was false. Conclude: the set of all (finite) palindromes is not first-order definable. q.e.d.

We will add some remarks concerning the above proposition.

Remark 7.2.5 (i) The logical structure of the proof showing that finite palindromes are not first-order definable is a proof by contradiction. This is the strategy that is often applied when Ehrenfeucht-Fraïssé games are used in finite model theory in order to show that a certain collection is not first-order definable. First, we assume that a property of a certain collection of structures was first-order definable and then we take two particular structures and show that there is a winning strategy for player II, i.e. that both structures make the defining formula for the property of the collection of structures true. If the two structures are appropriately chosen, one can deduce a contradiction.

(ii) We introduced the relation $\equiv_n$ in the game theoretical context as a relation between sequences of points that were played in an Ehrenfeucht-Fraïssé game. Because of the deeper connection between formulas of depth n that are true in a structure $\mathcal{A}$ if and only if these formulas are true in a corresponding structure $\mathcal{B}$ and the winning strategy of player II in an n-round Ehrenfeucht-Fraïssé game, we immediately have the following fact:[8] Assume $\mathcal{A}$ and $\mathcal{B}$ are two given structures. If player II has a winning strategy in every m-round Ehrenfeucht-Fraïssé game (for $m \leq n$), then every formula ϕ of depth at most n holds in $\mathcal{A}$ if and only if ϕ holds in $\mathcal{B}$. Therefore, $\mathcal{A}$ and $\mathcal{B}$ are equivalent structures with respect to formulas that have at most depth n. This was the reason why in model theory the notion of n-elementary equivalence was introduced: two models are n-elementary equivalent if they make the same formulas true that have depth at most n. Instead of introducing this model theoretic notion we introduced winning strategies of games. The accounts are equivalent.

(iii) Notice that in general we did not work with classical techniques of model theory. Instead of model theory, we applied the strong tool of Ehrenfeucht-Fraïseé games. Because of the fact that classical theorems cannot be generally applied in finite model theory (for example the compactness theorem, the Skolem-Löwenheim theorem, Robinson's theorem etc.), we need to use other techniques. The game theoretical approach is constitutive for the developed theory.[9]

We exemplified the limits of first-order definability using a very specific example (palindromes) and a relatively specific method (Ehrenfeucht-Fraïssé games). As a matter of fact, the method is nevertheless quite general. Using Ehrenfeucht-Fraïssé games one can show that the set of (finite) words of the form ww (where w is a word of a given finite alphabet), and the set of all finite connected graphs are not first-order definable, either. Better known examples of the non-definability of certain collections in first-order logic are: the set of natural numbers, the notions of finiteness and the property of a given structure to be infinite. Furthermore, the class of finite strict orderings

[8] Cf. Theorem 7.2.6.

[9] We mention that the problem of truth is not a problem of finite model theory, because truth can be defined in finite model theory. Truth becomes a problem if we work in infinite structures that have an elementary coding scheme.

of even cardinality is not first-order definable.[10] Last but not least, transitive closures of (finite) graphs are not first-order definable. Proper inductive and co-inductive definitions (on infinite structures) add an infinite number of additional examples for undefinability in first-order logic in a very general sense, because proper inductive (as well as coinductive) definitions are a second order concept.

It is worth mentioning that the problem of defining the set of certain finite structures can most times be reduced to the problem of defining precisely the collection of all finite objects (with a certain property). For example, one can easily define a collection of palindromes (over the alphabet {a,b}) using the following definition. Take the collection of all x, such that x satisfies:

$$\begin{array}{rl} & (x = a) \lor (x = b) \lor (x = aa) \lor (x = bb) \\ \lor & (x = aya \land y \in Pal) \lor (x = byb \land y \in Pal) \end{array}$$

The problem with this definition is that nothing prevents the generation of a collection of palindromes which contains palindromes that have infinite length. To get precisely the collection of all finite palindromes is the problem here. Additionally, to get precisely the collection of all finite and infinite palindromes is not first-order definable, either.

Concerning our second special feature - namely using Ehrenfeucht-Fraïssé games - we have to mention another possibility showing that certain notions are not first-order definable. This method uses essentially the compactness theorem.[11] Because it is well-known (and easy to prove) that the compactness theorem does not hold in second-order logic, we can quite easily derive contradictions from the assumption that a particular set is first-order definable.[12] We want to illustrate this by the undefinability of the property to be finite. In order to do this, we work with the concept of a theory. A theory Γ is a shortcut for a (not necessarily finite) set of first-order formulas.

Lemma 7.2.9 *There is no first-order theory Γ, such that Γ defines finiteness, i.e. for an arbitrarily given structure $\mathcal{A}$, it holds:*

$$\neg\exists\Gamma : \mathcal{A} \models \Gamma \iff |\mathcal{A}| \textit{ is finite} \qquad (\forall\phi \in \Gamma : \phi \textit{ is a first-order formula})$$

Proof: Assume towards a contradiction that finiteness was first-order definable. Then there exists a first-order theory Γ, such that

[10] The essential point here is the requirement of even cardinality and not the requirement of being a strict order relation.

[11] The compactness theorem can be stated as follows: Assume a structure $\mathcal{A}$ is given. Let Γ be a set of first-order formulas, such that for every finite $\Gamma' \subseteq \Gamma$ it holds: Γ' has a model. Then Γ has a model as well.

[12] Compare [BeDo83].

$$\mathcal{A} \models \Gamma \iff |\mathcal{A}| \text{ is finite.}$$

Now, consider for all $k \in \omega$ sentences of the following form:

$$\phi_k = \exists x_1 \exists x_2 .... \exists x_k : \bigwedge_{i<j} x_i \neq x_j \qquad (i, j \in \{1, 2, ..., k\})$$

Define a new theory Γ' as follows: $\Gamma' = \Gamma \cup \{\phi_k\}_{k<\omega}$. Take an arbitrary subset Γ_0 of $\Gamma \cup \{\phi_k\}_{k<\omega}$. Obviously, Γ_0 is consistent. Using the compactness theorem (we work in first-order logic, therefore the compactness theorem is valid), we find that Γ' is consistent, therefore Γ' has a model. Then: $\mathcal{A} \models \Gamma'$ which is a contradiction, because Γ' is infinite. q.e.d.

There is a variety of results concerning the limited power of first-order logic. Not only game theoretical means and the compactness theorem can be used. For example, Keisler proved that a property Φ of models is first-order definable if and only if Φ and its complement are closed under isomorphisms and ultraproducts.[13] We do not go into the details of these different ideas for proofs that first-order logic has limited expressive power. This section should simply show how games can be used to get results of the complexity of properties and collections of structures.

We mentioned some of the important properties of structures and collections of structures that are not first-order definable. The proofs for these claims are quite often very similar and can be formulated in various ways (dependent on the preferred approach), although they are in general not trivial. We stressed the game theoretical account here because it is a very general and quite intuitive approach towards definitional complexity. The purely model theoretic approach has the disadvantage to be quite abstract. Games have an intuitive basis and the principal idea of a winning strategy is simpler than the concept of n-elementary equivalence.

The reduced power of first-order logic defining relations and subsets of a given universe motivates directly different extensions of first-order logic. In fact, there are many different possibilities to enlarge the expressive power. One way is to go directly to full second-order logic or to a restricted kind of second-order logic like monadic second order logic (MSO). Other possibilities are to introduce types or additional operators (like in modal-logic, tense-logic, epistemic logic etc.). All these extensions allow an increase in the expressibility of the logical systems.

Informally, we will strengthen in the following sections (and in Chapter 8) first-order logic step by step by extending it using inductive definitions, coinductive definitions, and definitions of even higher complexity (in the analytical hierarchy), namely circular definitions. This will lead us to a better understanding of the definitional complexity.

[13] Cf. [BeDo83] or [ChKe73].

7.3 Inductive Definitions

The natural and most prominent example of an inductive definition is the definition of the natural numbers $\mathbb{N}$ via a successor operation $'$ and a designated element 0. Intuitively, an inductive definition specifies the smallest set in which the successor operation is closed and the whole set is generated from a bases case. The importance of inductive definitions in mathematics and the formal sciences cannot be underestimated. They are used in many different fields. We shall consider some examples of inductive definitions in order to get a flavor of the general idea.

Example 7.3.1 (i) The standard way to define the set of natural numbers $\mathbb{N}$ using a successor operation $'$ and a distinguished element 0 is specified in the following definition:

> $\mathbb{N}$ is the smallest set, such that the following relation holds:

$$(0 \in \mathbb{N}) \wedge (\forall y)(y \in \mathbb{N} \;\rightarrow\; y' \in \mathbb{N})$$

Standardly, 0 is interpreted as the empty set $\emptyset$. Whereas the above definition requires the existence of a structure and a successor function, the set theoretic version that can implicitly be found in most standard axiomatizations of ZFC can be formulated as follows:

> $\mathbb{N}$ is the smallest set, such that the following relation holds:

$$(\emptyset \in \mathbb{N}) \wedge (\forall y)(y \in \mathbb{N} \;\rightarrow\; y \cup \{y\} \in \mathbb{N})$$

In the above expression, $y \cup \{y\}$ corresponds to the successor operation $y' = y + 1$ in the first example. In other words, $y \cup \{y\}$ is the set theoretic code for the arithmetical expression $y + 1$. There are other possibilities as well. For example, $y' = y + 1 = \{y\}$ is another alternative for a set theoretical representation.

(ii) Consider again the collection of all finite palindromes Pal. We saw in Section 7.2 that the set of all finite palindromes over the alphabet $\{a, b\}$ is not first-order definable. What is needed in order to define Pal is the concept of an inductive definition. Consider again the finite alphabet $\{a, b\}$. The following inductive definition defines precisely the collection of all finite palindromes Pal over $\{a, b\}$.

> Pal is the smallest set, such that the following relation holds:

$$\begin{aligned}&(a \in Pal) \wedge (b \in Pal) \wedge (aa \in Pal) \wedge (bb \in Pal) \wedge \\ &\qquad\qquad \forall x : (x \in Pal \rightarrow (axa \in Pal \wedge bxb \in Pal))\end{aligned}$$

Like in the case of the natural numbers $\mathbb{N}$, we start the machinery with a base case (or several base cases as the example of the palindromes shows) and then we apply a production method to get new elements of Pal. The smallest set that is closed under the production process is the set of all finite palindromes over $\{a, b\}$.

(iii) Our last example for inductive definitions is well-known to logicians. The definition specifying the set of all well-formed expressions $Prop$ of propositional logic is essentially an inductive definition. To define the set $Prop$ we require the following:

> Assume a set of atomic proposition $AtProp = \{p, q, r, ...\}$ is given. Then, the collection $Prop$ of all well-formed formulas of propositional logic is the smallest set such that the following three conditions hold:
> (1) $\forall p : (p \in AtProp \to p \in Prop)$
> (2) $\forall \phi \forall \psi : (\phi \in Prop \wedge \psi \in Prop \ \to \ \phi \wedge \psi \in Prop)$
> (3) $\forall \phi : (\phi \in Prop \ \to \ \neg\phi \in Prop)$

The base cases are given by condition (1), namely that all atomic propositions are in $Prop$.[14] The inductive process operates on formulas that are built using logical connectives. It is worth mentioning that the requirement to choose the smallest set is essential for the whole consideration. Later we will see that it is possible to define a dual version of this idea where the extension of a concept is defined using the maximal fixed point of a certain operation. Although the ideas are quite similar, minimal and maximal fixed points have quite different properties in general.

Remark 7.3.2 We add some remarks concerning the fact that these definitions are second-order concepts. This is not obvious because the defining conditions are classical first-order conditions. The difficulty to find a first-order definition for these examples is not caused by the inductive process but by the requirement that the resulting set needs to be the smallest set satisfying some conditions. We formulate the definition of the natural numbers $\mathbb{N}$ once more again.

$$n \in \mathbb{N} \ \Leftrightarrow \ \forall S(\forall x(0 \in S \ \wedge \ (x \in S \ \to x' \in S)) \ \to \ n \in S)$$

In this definition, the second-order aspect becomes clearly visible. We have a second-order quantification in front of an arithmetical formula. As we will see later, every inductive definition is a Π^1_1 formula.

[14] Notice that in this example there are infinitely many base cases in general.

The above examples have a common feature: all definitions have certain (at most countably many) 'base cases' (and are therefore well-founded), whereas through a generation process more and more elements are added to the set which shall be defined. The generation process can be interpreted as a monotone operation via an operator Γ which takes, for example, (pairs of) formulas (of propositional logic) to formulas (of propositional logic), such that: $\Gamma(\phi, \psi) = \phi \wedge \psi$ or $\Gamma(\phi) = \neg\phi$. In Example 7.3.1(i), we have the following situation: a monotone operator Γ' maps numbers to numbers, such that, if $x \in \omega$, then $x + 1 \in \omega$, too. To get the set in question we have to generate the smallest set, which corresponds to the minimal fixed point of the monotone operator Γ. This connection establishes a possibility to give the probably most important characterization of inductive definitions. We take this characterization as the definition of an inductive definition. In other accounts, this characterization can be proven using an alternative definition of inductive definition.

Definition 7.3.1 *Given a language L, a definition D inductively defines a predicate P, if the extension of P is the minimal fixed point of a monotone operator Γ in the augmented language $L^+ = L \cup \{P\}$.*

In the monograph [Mo74], we can find important results of the theory of inductive definitions on abstract structures. In particular in this work, a general characterization theorem of the complexity of an inductive definition is proven relative to 'nice structures', namely so-called acceptable structures.[15] For every relation R on an acceptable structure the following equivalence holds:

$$R \text{ is a } \Pi^1_1 \text{ relation} \iff R \text{ is inductive}$$

In order to prove the above characterization, one uses standard game theoretic techniques. In the following, we shall consider some aspects of the game theory for inductive definitions. To do this, we have to modify our concept of Ehrenfeucht-Fraïssé games to a more general kind of game in order to be able to characterize infinite sets.[16] An important further restriction is the requirement that the infinite game is open (or closed). These notions will be defined below. Adopting the formalism in [Mo74], we begin with the introduction of a modified finite game which differs from the above Ehrenfeucht-Fraïssé games in an important respect. (The following ideas can be found in [Mo74], Chapter 4.)

Definition 7.3.2 *Assume R is a relation on a given set X and $\vec{Q} = Q_1 Q_2 \ldots Q_n$ is a sequence of quantifiers, i.e. $Q_i \in \{\exists, \forall\}$ for $i \in \{1, 2, \ldots, n\}$. The finite game $\Game(\vec{Q}, R)$ between player I and player II is defined as follows:*

[15] One can consider standard arithmetic as the natural example of an acceptable structure. Roughly speaking, an acceptable structure is a structure which contains an elementary coding scheme. The easiest example for an acceptable structure is standard arithmetic. Notice that in order to guarantee that the coding scheme is effective it is assumed that ever acceptable structure is countable.

[16] Notice that Ehrenfeucht-Fraïssé games are essentially finite games. We would like to work with infinite games as well.

- *Player I and player II play elements of X. If $Q_i = \exists$ then player I plays $x_i \in X$ and if $Q_i = \forall$ then player II plays $x_i \in X$.*

- *Player I wins $\Game(\vec{Q}, R)$ if $R(x_1, x_2, \ldots, x_n)$ holds and player II wins $\Game(\vec{Q}, R)$ if $\neg R(x_1, x_2, \ldots, x_n)$ holds.*

Remark 7.3.3 (i) Recall the definition of Ehrenfeucht-Fraïssé games above: two players play in two structures alternately. Definition 7.3.2 does not use structures at all. Simply prefixes of formulas determine a particular game. Furthermore, players I and II do not play points of a structure but rather subsets of the given set. Moreover, in the game the players do not alternately play sets, but rather the linear sequence of moves of the game is determined by the properties of $\vec{Q}$. Finally, we do not want to find a partial isomorphism between two given structures in a game, but we want to model a given relation R using sequences of quantifiers. In fact, the games considered her are quite different from Ehrenfeucht-Fraïssé games.

(ii) Definition 7.3.2 is the finite version of a game in infinite structures. The infinite version of games as defined in Definition 7.3.2 are essentially used to characterize inductive and coinductive definitions. The finite Ehrenfeucht-Fraïssé games we saw in Section 7.2 were used to show that a certain property of a certain relation is not first-order definable. Hence, the general idea of these games is rather different.

What can be said about winning strategies in the games as defined in Definition 7.3.2? We define a winning strategy for player I as follows:

Definition 7.3.3 *(i) Player I has a winning strategy in $\Game(\vec{Q}, R)$, if there exists a set of functions $S = \{f_i \mid i \in \omega \ \wedge \ Q_i = \exists \ \wedge \ arity(f_i) = i - 1\}$ such that player I wins every run of $\Game(\vec{Q}, R)$ whenever player I follows S.*

(ii) Player II has a winning strategy in $\Game(\vec{Q}, R)$, if there exists a set of functions $S = \{f_i \mid i \in \omega \ \wedge \ Q_i = \forall \ \wedge \ arity(f_i) = i - 1\}$, such that player II wins every run of $\Game(\vec{Q}, R)$ whenever player II follows S.

Notice that we separate the definitions of winning strategies for both players. For finite games, it is clear that if player I has no winning strategy, then player II must have a winning strategy. On the other hand, if player II has no winning strategy, then player I must have a winning strategy. Furthermore, in every finite game, either player I or player II has a winning strategy. This determinacy condition does not necessarily hold for infinite games as defined below. We will add some more remarks concerning this important determinacy aspect later.

We saw above that the set of (finite) palindromes over an alphabet with two elements is not first-order definable. Other examples are: the set of finite words of the form ww, the set of natural numbers, arithmetical addition, the set of all countable graphs which are connected, the set of all countable graphs such that there are no two points that have infinite distance.[17] All these relations can be defined inductively. For our context, that means the following: we cannot hope to find a finite string of quantifiers $\vec{Q} = Q_1 Q_2 \dots Q_n$ and a game $\Game(\vec{Q}, R)$, such that $\Game(\vec{Q}, R)$ can be (in any sense) associated with an inductively defined relation. We need to extend the present account in order to work with infinite prefixes of quantifiers.

Consider an infinite string of quantifiers $\vec{Q} = Q_1 Q_2 \dots$ and an infinite relation R over a set X. We use the following shortcut in order to simply the formalization of the theory:

$$\vec{Q}(R(\vec{x})) \;\Leftrightarrow\; \{Q_1 x_1 Q_2 x_2 \dots\} R(x_1, x_2, \dots)$$

The definitions for our game remain the same, except that player I has a winning strategy for $\Game(\vec{Q}, R)$ if player I wins every run of the infinite game $\Game(\vec{Q}, R)$, i.e. if $R(x_1, x_2, ...)$. Player II has a winning strategy if player II wins every run of the infinite game, i.e. if $\neg R(x_1, x_2, ...)$. Because these concepts are important, we formulate these concepts in a separate definition.

Definition 7.3.4 *Assume R is an infinite relation on a given set X and $\vec{Q} = Q_1 Q_2 \dots$ is an infinite sequence of quantifiers, i.e. $Q_i \in \{\exists, \forall\}$ for $i \in \{1, 2, \dots\}$. The infinite game $\Game(\vec{Q}, R)$ between player I and player II is defined as follows:*

- *Player I and player II play elements of X. If $Q_i = \exists$ then player I plays $x_i \in X$ and if $Q_i = \forall$ then player II plays $x_i \in X$.*
- *Player I wins $\Game(\vec{Q}, R)$ if $R(x_1, x_2, \dots)$ holds and player II wins $\Game(\vec{Q}, R)$ if $\neg R(x_1, x_2, \dots)$ holds.*

In order to formulate the relationship between winning strategies of infinite games and inductive definitions, we need finer distinctions of infinite relations. The following definition distinguishes two different kinds of relations.

Definition 7.3.5 *(i) An infinite relation $R(x_1, x_2, \dots)$ is called open, if there are relations $R_1(x_1)$, $R_2(x_1, x_2)$, $R_3(x_1, x_2, x_3)$,..., such that it holds:*

$$R(x_1, x_2, \dots) \iff R_1(x_1) \vee R_2(x_1, x_2) \vee R_3(x_1, x_2, x_3) \vee \dots$$

(ii) An infinite relation $R(x_1, x_2, \dots)$ is called closed, if there are relations $R_1(x_1)$, $R_2(x_1, x_2)$, $R_3(x_1, x_2, x_3)$, ..., such that it holds:

$$R(x_1, x_2, \dots) \iff R_1(x_1) \wedge R_2(x_1, x_2) \wedge R_3(x_1, x_2, x_3) \wedge \dots$$

[17] Notice that the last example is not equivalent to the set of all finite graphs, if finite graph means 'there is a maximal path of length $n \in \omega$ in the graph'.

An important point concerning the infinite case is that not everything generalizes from the finite case to the infinite one. The difficulty is the (already mentioned) determinacy of games. Whereas in the finite case, either player I or player II has a winning strategy (this is a trivial non-logical result), this is no longer true in general for infinite games. An example of an infinite game that is not determined can be found in [Je78], p.551. The famous insight in [GaSt53] was the fact that in ZFC one can prove that every infinite game $\Game(\vec{Q}, R)$ between player I and player II is determined, provided that R is an open or closed relation. In other words: in every game $\Game(\vec{Q}, R)$ where R is either open or closed (or both), either player I or player II has a winning strategy. This fact is called the Gale-Stewart theorem.[18] There are different versions of the Gale-Stewart theorem. Here, we adopt the version that is formulated in [Mo74]:

Theorem 7.3.6 *Assume A is a non-empty set and $R \subseteq A_\omega$ is an open or closed relation. Let $\vec{Q} = Q_1Q_2\ldots$ be an infinite string of quantifiers. Then, player I or player II has winning strategy in $\Game(\vec{Q}, R)$.*

Proof: Compare [Mo74], 4A.1, pp.56/57. q.e.d.

We add some further remarks, comments, and implications of the Gale-Stewart theorem.

Remark 7.3.4 (i) In general, the Gale-Stewart theorem gives us a lower bound of all sets that can be associated with infinite games. We know from Theorem 7.3.6 that at least the open and closed sets are possible payoffs of determined infinite games. Theorem 7.3.6 does not tell us anything about arbitrary second-order relations. In particular, it is not clear whether arbitrary relations can be associated with infinite games.

(ii) Determinacy of games can be formulated in other terms as the following formula shows:

$$\forall \Game(\vec{Q}, R) : \neg\vec{Q}(R(\vec{x})) \iff \vec{Q}^{\smile}(\neg R(\vec{x}))$$

In this context, $\vec{Q}^{\smile}$ means the 'dual' infinite prefix of $\vec{Q}$. Formally, we can define $\vec{Q}^{\smile}$ as follows:

$$\{Q_1x_1Q_2x_2...\}^{\smile}(R(x_1, x_2, ...)) \iff Q_1^{\smile}x_1Q_2^{\smile}x_2...(\neg R(x_1, x_2, ...))$$

where we take $\exists^{\smile} = \forall$ and $\forall^{\smile} = \exists$. Determinacy means that a negation in front of the formula $R(x_1, x_2, ...)$ can be extracted. We know this situation from the finite case in ordinary first-order logic as the following simple equivalence shows:

$$\neg\exists x\forall y : \phi(x, y) \iff \forall x\exists y : \neg\phi(x, y)$$

[18] Cf. [GaSt53].

Determinacy means that a negation in front of the prefix of the formula can be incorporated into the formula, such that the negation can be placed behind the prefix and the quantifiers are substituted by their duals without changing the meaning of the whole expression.

(iii) An example of a relation which is neither open nor closed and where the corresponding game is not determined can be found in [Je78], p. 551. The example is crucially based on a diagonal argument and essentially uses the axiom of choice. It is still not known whether there exists a constructive example of a game that is not determined (i.e. an example that does not use the axiom of choice).

(iv) An important discussion in set theory is the question whether an additional axiom to ZFC can guarantee determinacy. It turns out that full determinacy (standardly called the axiom AD, interpreted as the statement that every infinite game is determined) is inconsistent with the axiom of choice (this is an immediate consequence of (iii)), whereas a weaker form of the axiom of choice is implied by AD. Therefore, a weaker form of determinacy seems to be more reliable for mathematical purposes. An important candidate for weak determinacy is the axiom which states the determinacy of the projective hierarchy (usually called PD). With this axiom, every second-order relation can be associated with an infinite game. This new axiom added to ZFC yields a consistent theory (especially it is consistent with the axiom of choice). We will return to this point, below.

The important result concerning inductive definitions is the theorem that associates Π^1_1-relations with relations that are inductively defined, provided we work in acceptable structures. We mention this theorem of the theory of inductive definitions without a proof.

Theorem 7.3.7 *A relation R on an acceptable structure is Π^1_1 iff R is inductive.*

Proof: Compare [Mo74], 1D.3, p.20 (for the ”$\Leftarrow$” direction), and 8A.1, pp.132-134 (for the ”$\Rightarrow$” direction). q.e.d.

Remark 7.3.5 The restriction of Theorem 7.3.7 to acceptable structures is essential because only in this case there exists an appropriate (elementary) coding scheme. Acceptable structures are essentially arithmetical-like structures that allow an appropriate coding of formulas and relations. The idea of coding formulas and relations in acceptable structures is quite similar to the coding scheme used in recursion theory. One can take standard arithmetic as the archetypical example. For more information concerning acceptable structures the reader is referred to [Mo74].

The following theorem gives us the connection between open relations and inductive relations. Furthermore, it provides a further characterization of inductive relations.

Theorem 7.3.8 *A relation R on an acceptable structure is open if and only if R is inductive.*

Proof: Straightforward consequence of the proof of Theorem 7.3.7. Compare [Mo74]. q.e.d.

Optimal would be a result that associates inductive definitions with a straightforward generalization of Ehrenfeucht-Fraïssé games we examined in Section 7.2. The infinite games used in this section to model inductive definitions are not comparable with Ehrenfeucht-Fraïssé games because of the following differences:

(i) Ehrenfeucht-Fraïssé games are finite, whereas the described games in this section are infinite.

(ii) Ehrenfeucht-Fraïssé games are defined on two (usually finite) structures, whereas the games in this section are defined on one (generally infinite) structure.

(iii) Ehrenfeucht-Fraïssé games are determined by the moves on structures, whereas the games in this section are determined by the prefixes of given formulas.

It turns out that an association of Ehrenfeucht-Fraïssé games with infinite games is problematic. A naive infinite extension of Ehrenfeucht-Fraïssé games from finite games to infinite ones makes it impossible to distinguish different structures of cardinality $\aleph_0$ as we will see below. Precisely this would be necessary. We will examine in the following the problems of a naive extension of Ehrenfeucht-Fraïssé games to the infinite.

In order to do this, consider infinite Ehrenfeucht-Fraïssé games that are specified as follows: Definition 7.2.2 remains the same, except that we play now infinite games. Player II has a winning strategy in the infinite Ehrenfeucht-Fraïssé game if player II can respond to every move of player I in an arbitrary infinite sequence of moves, such that there is a partial isomorphism between the structures.[19] Notice that 'having a winning strategy in the infinite game' is not equivalent to 'having a winning strategy in every n-round game'. A counterexample can be given as follows: let $\mathcal{G}$ and $\mathcal{H}$ be two given structures where $\mathcal{G}$ is the set of all n-circles for $n \in \omega$, and $\mathcal{H}$ is the collection of all n-circles plus a copy of the integers. We assume that the relevant relation on the integers is 'x is the neighbor of y' (connectedness). Whereas player II has a winning strategy in every n-round game (because player II can choose 'big enough circles in every finite game'), player II has no winning strategy in the

[19] Notice that a partial isomorphism needs not to be a function with a finite domain.

infinite game. The latter is true because player II cannot choose an infinite circle in structure $\mathcal{G}$, simply because there is no infinite circle in $\mathcal{G}$.

The next propositions specify the relation between infinite Ehrenfeucht-Fraïssé games and countable structures. It turns out that if player II has a winning strategy in the infinite Ehrenfeucht-Fraïsé game, then the two structures are isomorphic.

Proposition 7.3.9 *If $\mathcal{A}$ and $\mathcal{B}$ are structures with $|\mathcal{A}| \leq \aleph_0$ and $|\mathcal{B}| \leq \aleph_0$. Then it holds: player II has winning strategy in every infinite Ehrenfeucht-Fraïssé game if and only if $\mathcal{A} \cong \mathcal{B}$.*

Proof: "$\Rightarrow$" Assume that player II has a winning strategy in every infinite Ehrenfeucht-Fraïssé game. We have to show that $|\mathcal{A}| = |\mathcal{B}|$ and that there is a structure preserving function (homomorphism) $f : |\mathcal{A}| \longrightarrow |\mathcal{B}|$. Because A is at most countable, we can fix an enumeration of the elements of $\mathcal{A}$. (It is not problematic, if there is no recursive way to do this.) Then, consider an Ehrenfeucht-Fraïssé game between player I and player II where player I plays only in structure $\mathcal{A}$ and player II plays only in structure $\mathcal{B}$. Furthermore, assume that in this game player I enumerates all elements of $\mathcal{A}$. Because player II has a winning strategy in the infinite game, there is a partial isomorphism $f : A \longrightarrow B$, such that f is defined on the entire domain A. In other words, A is isomorphically embedded in B. Therefore, it holds: $A \preceq B$, i.e. A is injectively embedded into B.
Now we play the reverse game. Player I plays in structure $\mathcal{B}$ and player II plays in structure $\mathcal{A}$. Because player II has a winning strategy in this game, we get $B \preceq A$. On a set theoretical level we can deduce (using the Schröder-Bernstein Theorem): $A \cong B$. On a model theoretical level, we get $\mathcal{A} \cong \mathcal{B}$ because player II has a winning strategy in the infinite game, i.e. every game determines a partial isomorphism between $\mathcal{A}$ and $\mathcal{B}$.

"$\Leftarrow$" If $\mathcal{A} \cong \mathcal{B}$, then there is an isomorphism mapping $\mathcal{A}$ into $\mathcal{B}$. Take this isomorphism as a device for player II. Then, player II has a winning strategy in every infinite game. q.e.d.

As a further proposition, we can state the following fact which is the pendant to Theorem 7.2.6.

Proposition 7.3.10 *Let $\mathcal{A}$ and $\mathcal{B}$ be structures with $|\mathcal{A}| \leq \aleph_0$ and $|\mathcal{B}| \leq \aleph_0$. Assume player II has a winning strategy in every infinite Ehrenfeucht-Fraïssé game where player I plays in $\mathcal{A}$ and player II plays in structure $\mathcal{B}$. Then it holds:*

$$\forall \phi : (\mathcal{A} \models \phi \ \Rightarrow \ \mathcal{B} \models \phi)$$

Proof: We have to show that there is an isomorphic embedding f mapping $\mathcal{A}$ into $\mathcal{B}$. Consider an arbitrary enumeration of the elements of A. Assume further that a game is specified where player I plays a_1 in the first round, a_2 in the second, and so on. Because player II has a winning strategy in the infinite game, player II plays b_1 in round 1, b_2 in round 2, and so on. Player II has a winning strategy (assumption), therefore $f : a_i \longmapsto b_i$ is a partial isomorphism. Conclude: f embeds $\mathcal{A}$ into $\mathcal{B}$ isomorphically. q.e.d.

Remark 7.3.6 (i) Notice that it is not necessary to require determinacy of the games. Our account is different: we presuppose that player II wins every infinite game, i.e. the games are determined.

(ii) Kripke's fixed point theory is essentially based on inductive definitions. Hence, the complexity of his theory is Π^1_1. The construction of minimal fixed points fits nicely into the presented account. We will reconsider the properties of Kripke's theory in more detail in Chapter 16.

(iii) As far as the author knows, it is an open question whether it is possible to generalize Ehrenfeucht-Fraïssé games to games defined on infinite structures, such that we get a reasonable correspondence to Moschovakis' open (or closed) games. To connect open games with the Ehrenfeucht-Fraïssé account, it seems reasonable to assume that the payoffs of Ehrenfeucht-Fraïssé games are codes of partial isomorphisms, such that the set of all these codes is itself an open subset of the natural numbers. But this is not worked out yet.

As we saw in the above considerations, the natural dual construction of inductive definitions are coinductive definitions. We will add some remarks concerning coinductive definitions in the next section. Most of the described properties of coinductive definitions are simply the dual of the properties of this section.

7.4 Coinductive Definitions

From a category theoretic perspective, coinductive definitions are dual constructions of inductive definitions. We saw several alternative characterizations of inductive definitions in the last section. Coinductive definitions can be interpreted in a very analogous way. The following definitions make the idea of a coinductive definition precise.

Definition 7.4.1 *Given a language L a definition D coinductively defines a predicate P, if the extension of P is determined by the maximal fixed point of a monotone operator Γ in the augmented language $L^+ = L \cup \{P\}$.*

Remark 7.4.1 Quite often, it happens that maximal fixed points are coextensional with minimal fixed points. For example, consider the domain $(a, b)^*$, i.e. the set of all finite words of the alphabet $\{a, b\}$. Determine the largest set $PAL \subseteq (a, b)^*$, such that it holds:

$$\begin{aligned}&(a \in PAL) \wedge (b \in PAL) \wedge (aa \in PAL) \wedge (bb \in PAL) \wedge \\ &\qquad\qquad \forall x \in (a,b)^* : (x \in PAL \to (axa \in PAL \wedge bxb \in PAL))\end{aligned}$$

We find that this set PAL is equal to Pal, the set determined by the minimal fixed point. Notice that in this example we have already established an upper bound for the length of words, namely we choose possible candidates from the collection of all finite words. The second-order concept is hidden in the assumption that $PAL \subseteq (a,b)^*$. Differences between the maximal and the minimal fixed point arises if we take (countably) infinite words (i.e. infinite sequences of the letters of the alphabet) into account. Then, we get as the minimal fixed point the set

$$Pal = \{w \in (a, b^*) \mid w \text{ is a finite palindrome}\}$$

The maximal fixed point PAL is the collection of all finite and (countably) infinite palindromes. Hence, the two fixed points are not coextensional.

We will mention some examples of coinductive definitions. Later, we will see that these examples can be used in the theory of hypersets as well.

Example 7.4.2 (i) Assume a set of atomic proposition $AtProp = \{p, q, r, \dots\}$ is given. The collection of all propositions $PROP$ is defined as the largest set, such that it holds:

$$\begin{aligned}\forall \phi \in PROP : \quad & \phi \in AtProp \\ \vee \quad & (\phi = \psi \wedge \chi \text{ and } \psi \in PROP \text{ and } \chi \in PROP) \\ \vee \quad & (\phi = \neg\psi \text{ and } \psi \in PROP)\end{aligned}$$

The above definition is a typical coinductive definition. Notice that nothing prevents the existence of infinite formulas. In fact, the existence of propositional formulas of infinite length in $PROP$ guarantees that $PROP$ and $Prop$ are not equal. Whereas the requirement in the inductive case to take the smallest set avoids formulas of infinite length, this is not true for the largest collection.

(ii) Consider the following coinductive definition. ORD is the largest collection such that the following condition holds (where $'$ is the successor function):

$$\forall \alpha \in ORD : \ \alpha = 0 \ \vee \ (\alpha = \beta' \text{ and } \beta \in ORD)$$

The above definition specifies the class of all ordinals. This class ORD is the largest collection closed under the successor function. Clearly, by a restriction of a certain cardinal number we can also define an initial segment of the ordinals using a coinductive definition.

Similarly to inductive definitions, coinductive definitions can be identified as collections of a certain complexity class of the projective hierarchy. It turns out that coinductively defined predicates specify negations of inductively defined predicates. This is stated in the following fact.

Fact 7.4.2 *A relation R on an acceptable structure is Σ^1_1 if and only if R is coinductive.*

Proof: Dual of Theorem 7.3.7. q.e.d.

The second-order quantification in coinductive definitions comes in when the existence of a largest collection satisfying certain properties is claimed. It is a quite similar feature as in the inductive case. In general, it seems to be the case that inductive definitions are more prominent in mathematical discourse. The reason for this is probably the fact that we most often use (infinite) collections of finite objects in mathematics. Coinductive definitions specify quite often (infinite) collections of infinite objects.

Finally, we can establish the connection between coinductive definitions and closed games. This comes down to the equivalence of coinductive relations and closed relations (again as the dual statement of the inductive case). We state this as a fact without proof.

Fact 7.4.3 *A relation R on an acceptable structure is closed if and only if R is coinductively defined.*

Proof: Dual of Theorem 7.3.8. q.e.d.

Remark 7.4.3 (i) We have the following correspondences: coinductive definitions are precisely the Σ^1_1 relations (on acceptable structures). Σ^1_1 relations are equivalent to the extension of a maximal fixed point of a monotone operator Γ. Finally, it holds that coinductive relations are equivalent to closed relations.

(ii) Game theoretically interpreted, the described results mean that coinductive relations can be associated with closed relations, i.e. they can be represented as an infinite disjunction of finite relations. If R is an inductive relation, then player II has a winning strategy and on the other hand, if R is a coinductive relation, then player I has a winning strategy in the corresponding infinite game. Notice that for relations R which are inductively definable as well as coinductively definable, the winning strategies depend on the representation of R. We know that a Δ^1_1 relation R can be represented either as a relation of the form $\forall X : \phi(X)$, or as a relation of the form $\exists X : \psi(X)$. The same is true if R is written as an infinite sequence of first-order quantifiers. Dependent on the representation, either player I has winning strategy or player II has a winning strategy. The relation R itself is closed and open.

We finish this section with these remarks concerning coinductive definitions. Much more could be said concerning this topic. We do not develop the theory of coinductive definitions further, because coinductive relations are not the focus of this part of this work. In Part IV, we will consider coinductive definitions again. In the next chapter, we will develop revision theories as a further extension of classical definition theory that goes beyond inductive and coinductive definitions. The added complexity comes in because fully circular definitions are allowed in revision theories.

7.5 History

The topics in this chapter possess a quality of folklore. Ehrenfeucht-Fraïssé games are based on the papers [Fr54] (algebraic foundation) and [Eh61]. In the latter work, Ehrenfeucht introduced a game theoretical characterization of partial isomorphisms. On these ideas, the complete theory of finite models is based. A good introduction to finite model theory is [EbFl95]. Other sources of information are monographs in complexity theory (usually in the field of computer science) like [Im99]. Inductive definitions are a very old idea, although an algebraic theory of inductive definitions (especially including its complexity theoretic properties) is a rather new development. The origins of a theory of inductive definitions grew out of a number of papers Stephen C. Kleene published around the year 1955. Most prominently in this context, we have to mention Kleene's work [Kl55]. This paper can count as the basis for all further developments done by Moschovakis, Spector, and Barwise. The standard monograph of inductive definitions is [Mo74]. Important further results can be found in [Bar75]. The game theoretic perspective seems to become more and more important in different fields of mathematics and computer science.

Chapter 8

The Gupta-Belnap Systems

We examined in the last chapter a possibility to determine whether a relation is first-order definable. Additionally, we saw two extensions of first-order definability in terms of inductive and coinductive definitions. Now, inductive definitions do not give us full second-order logic and coinductive definitions either. We maximally get the complexity classes $\Sigma^1_1 \cup \Pi^1_1$, but it is impossible to define a proper Δ^1_2 relation, not to mention relations of even higher complexity. In order to get stronger theories, we will extend the expressive power of structures by circular definitions in the sense of [GuBe93]. The resulting theories are called Gupta-Belnap systems (sometimes we will use the general notion revision theories without a further specification). We will see that using revision theories we get a further possibility to strengthen given structures. An additional aspect of these theories is that they have more expressive power than inductive definitions (or coinductive definitions).

8.1 Circular Definitions and Revision Theories

Gupta-Belnap revision theories can be understood as a method to increase the expressive power of a given structure. Originally, Gupta-Belnap systems were developed to express and to give a reasonable (i.e. appropriate and consistent) interpretation of circular definitions. First, we shall examine some basic properties and results of this theory, and we shall add some examples for a better understanding of the new definitions. Second, we will prove that a certain semantical system (namely $\mathbf{S}^*$) has complexity Π^1_2. This result is quite important because it tells us that the system $\mathbf{S}^*$ is not recursively axiomatizable, i.e. $\mathbf{S}^*$ is a proper second-order system. The Gupta-Belnap account examines infinitely many semantical systems, in addition to the systems $\mathbf{S}^*$ and $\mathbf{S}^\#$ we are interested in. We will not consider certain finite (first-order) systems $\mathbf{S}_n$ (introduced also by Gupta and Belnap), because the properties of these systems are summarized in length in [GuBe93]. For further information of these systems as well as the definition of the corresponding calculi, the reader is referred to [GuBe93].

We are assuming a given first-order language L, and we extend L to L^+ via one or more (but finitely many) circular definitions. The ground model $\mathfrak{M}$ is considered a classical model, i.e. a model in which an evaluation is based on classical logic (not on a special kind of many-valued logic, or in general non-classical logics, although this would not change important parts of the theory) and classical model theory. As in Kripke's account, the ground model will give us an interpretation of the unproblematic sentences. The interesting part of Gupta-Belnap systems is the extension of the ground model in order to find an interpretation of the problematic circular definitions.

Circular definitions can be understood as definitions, where the definiendum occurs in the definiens. Notice that this is impossible in classical definition theories. Such a definition would be ill-formed according to the classical understanding, because in classical model theory one can deduce contradictions from such an unrestricted definition.[1] The motivation for the development of revision theories was originally motivated by the task to find a truth theory for formal as well as natural languages. Therefore, revision theories were embedded in a practical context, dealing with pathological sentences and paradoxes. We will restrict our consideration for the moment to the mathematical entities that are involved, because in this chapter we want to examine primarily the mathematical framework and its properties rather than possible applications.

Let us begin with an example. Assume we want to define a new (circular) predicate G. The following definition would be an example for such a circular definition (provided an appropriate domain D is given, for example, the class of all ordinals):

$$\forall x \in D : G(x) \iff (x = 0) \vee \exists y(G(y) \wedge x = y') \wedge \neg\exists z(G(z) \wedge z' = 0)$$

Intuitively, the extension of G should give us the natural numbers $\mathbb{N}$, i.e. we wish to define the natural numbers with the above circular definition. In order to reach this goal, we need an appropriate semantical system to evaluate the above definition. Notice that for classical model theory there is no way to evaluate the extension of G. Comparing the above definition with the corresponding inductive definition in Example 7.3.1(i), we see that we do not need to construct the smallest collection of elements of the domain. The above circular definition specifies precisely the natural numbers. Gupta and Belnap proposed many different semantical systems that can be used to define validity concerning circular definitions. We will mostly restrict our consideration to the semantical system $\mathbf{S}^*$ (and implicitly to the system $\mathbf{S}^\#$ as well) as a special but most interesting case.

The heart of revision theories is in a certain sense the concept of a revision sequences: under the assumption that h is a hypothesis for an extension of a predicate P, we apply infinitely often a given revision rule ρ and check the

[1] It is important to notice that the intended circular definitions are not implicit definitions in classical model theory. By the theorem of Beth one can show that every implicit definition can be represented as an explicit definition (cf. [ChKe73]). Hence, implicit definitions do not strengthen the given theory. As we will see this is quite different using Gupta-Belnap systems.

outcome of this operation. ρ itself is induced by circular definitions specifying the successor step in the revision process. We can understand revision sequences as approximations of hypotheses to better and better extensions. Intuitively, if the Liar sentence is true in our initial hypothesis h, then the next revision makes the Liar sentence false and the following revision makes it true and so on.

We need to define the following concept: an element d of domain D is stably x in a given sequence S, where a sequence S of hypotheses is a sequence of functions $f : D \longrightarrow X$.

Definition 8.1.1 *(i) Assume S is a sequence of hypotheses and D and X are non-empty sets. We say $d \in D$ is stably $x \in X$ if and only if the following condition holds:*

$$\exists\alpha \in ORD : (\alpha < length(S) \wedge \forall\beta \in ORD : \\ (\alpha \leq \beta \leq length(S) \rightarrow S_\beta(d) = x))$$

(ii) An infinite sequence S of hypotheses $h \in X^D$ is a revision sequence for the revision rule $\rho : X^D \longrightarrow X^D$ (where X and D are non-empty sets) if and only if conditions (a) and (b) hold:

(a) $S_{\alpha+1} = \rho(S_\alpha)$

(b) If α is a limit ordinal, then the following condition holds:

$$(\forall d \in D)(\forall x \in X) : (d \textit{ is stably } x \in S \restriction \alpha) \rightarrow S_\alpha(d) = x$$

Remark 8.1.1 (i) Notice: It is not required that S is finite. Actually, S can be arbitrarily long (for example S can be of length ORD). Our revision rule $\rho : X^D \longrightarrow X^D$ works in the general case where X and D are arbitrary sets. An easy example for a revision rule would be the following: take $D = \{\lambda\}$ (the Liar sentence), $X = \{t, f\}$ (truth values) and define the revision rule as follows: if $\rho^\alpha(\lambda) = t$, then $\rho(\rho^\alpha(\lambda)) = f$ and if $\rho^\alpha(\lambda) = f$, then $\rho^\alpha(\rho(\lambda)) = t$. Clearly, we have the well-known behavior of the Liar sentence.

(ii) The idea of Definition 8.1.1(ii) is that at a successor stage the application of the given revision rule ρ yields a new 'approximation' of the extension of the new predicate. Because we can start with arbitrary choices for a hypothesis (at stage 0), there are in general infinitely many possible revision sequences for one revision rule. Whereas at a successor stage α there are no other choices for a reasonable definition of the $\alpha + 1^{st}$ revision, the limit case is more problematic: what is a good extension for the definiendum in the limit case with respect to the unstable elements? The account in [GuBe93] allows arbitrary choices for the unstable elements. Only the stable elements $d \in D$ remain fixed. This treatment is usually called the Gupta-Belnap limit rule. Another possibility is to fix an interpretation of the unstable elements at the limit stage in advance. This limit rule is called the Herzberger rule (cf. [He82a, He82b]). Finally, one can always choose the interpretation of the initial hypothesis that results in the so-called constant limit rule (cf. [Gu82]). In a certain sense, Definition 8.1.1 is the most liberal one, because there is no a priori reason to prefer one version or the other.[2] Definition 8.1.1 has an additional advantage: it is the most general one, and therefore the most flexible one because it determines only the stable elements and lets the unstable elements unspecified at the limit stage.

(iii) Dependent on the semantical system, sometimes it is completely sufficient to consider revision sequences S that have finite length or that are at least bound by an ordinal: for example, $length(S) \leq \lambda$ for a limit ordinal λ. Unfortunately, we will work in relatively strong semantical systems where it is necessary that most times the considered revision sequences S have length $length(S) = ORD$. That implies that each revision sequence S itself is no longer a set but a proper class. That could cause problems in the case where there was no possibility to reduce the whole considerations to a set theoretical level. Fortunately, such a transition is possible. The standard way in order to reduce the complexity of these objects can be achieved by applying a Löwenheim-Skolem-style argument.[3]

In Definition 8.1.1, we assumed that a revision rule $\rho : X^D \longrightarrow X^D$ is given. The intuition how this revision rule is generated can be explained as follows: given a circular definition of a predicate P, this definition induces a revision rule ρ for the successor stages of the evaluation. For example, the Liar sentence changes its truth value after every revision and the Liar circle of length 2 changes its truth value after every second revision. Hence, a circular definition induces a revision rule ρ and this revision rule is used to specify a revision sequence S. Later, we will see that revision sequences will be used in order to develop a semantics for circular definitions. In order to simplify the notation, we avoid to write $S_{D,\rho}$ when referring to revision sequence S, that is induced by the circular definition D and revision rule ρ. It is clear from the context which revision rule must be applied.

Whereas Definition 8.1.1(i) gives an account for elements of the domain which become stable in the revision process, one could imagine the following situation. A particular element does not fix its interpretation after a certain infinite number of revisions, but fixes its interpretation after a certain infinite number of revisions, plus an additional finite number of further revisions. In other words: it is allowed that in a sequence with cofinality ω, finitely many fluctuations arise till the next limit ordinal is reached. This possibility is captured in Gupta's concept of near stability. In the next definition, we will state the concept of near stability. Although this concept is important for certain semantical systems, we will not examine this concept in full detail in this work.

[2] At the early developments of revision theories there was a long discussion concerning the limit rules. Obviously, it is hard to argue a priori for or against one of these rules. In a particular example, it is possible to choose one, in order to get better results. That was originally the argument of Herzberger in order to promote his choice of the Herzberger limit rule. We come back to a discussion of these rules later.

[3] Cf. Theorem 8.2.2. This fact was proven by McGee (cf. [Mc91]).

The reason for this is the fact that our focus is directed to the semantical system $\mathbf{S}^*$. The semantical system that includes near stability is the system $\mathbf{S}^\#$, but this system has quite similar properties to $\mathbf{S}^*$ (at least from a global perspective).

Definition 8.1.2 *Assume S is a sequence of hypotheses and D and X are non-empty sets. We say $d \in D$ is nearly stable at $x \in X$ iff the following condition holds:*

$$\exists \alpha \in ORD : (\alpha < length(S) \wedge \forall \beta \in ORD : \\ (\alpha \leq \beta < length(S) \rightarrow (\exists n \in \omega : (\forall m \geq n : S_{\beta+m}(d) = x))))$$

In the next definition, we introduce two important concepts. The intention is to define the property of a hypothesis to occur over and over again in a certain revision sequence.

Definition 8.1.3 *(i) Assume X and D are non-empty sets, $\rho : X^D \longrightarrow X^D$ is a revision rule (induced by a definition), and S is a revision sequence for ρ. A hypothesis $h \in X^D$ is called cofinal in S if and only if the following condition holds:*

$$(\forall \alpha \in ORD)(\exists \beta \in ORD) : \alpha < \beta \ \wedge \ S_\beta = h$$

(ii) A hypothesis $h \in X^D$ is called recurring for ρ if and only if there is a revision sequence S of length ORD, such that h is cofinal in S.

(iii) A hypothesis $h \in X^D$ is called α-reflexive for a revision-rule ρ if and only if there is a revision sequence S for ρ, such that $S_0 = S_\alpha = h$. A hypothesis $h \in X^D$ is called reflexive for ρ if and only if there is an ordinal α, such that h is α-reflexive.

Definition 8.1.3 makes it possible to analyze revision sequences. Although the intuitive idea of a revision sequence is easy to understand, sometimes revision sequences have quite surprising properties. We will examine some of these properties in several examples later. Some remarks concerning Definition 8.1.3 should already be added here.

Remark 8.1.2 (i) The general idea of a cofinal hypothesis is the following one: cofinal hypotheses occur over and over again in a revision sequence S. In other words, if a cofinal hypothesis h occurs somewhere in a revision sequence S, then we can find h again after at most γ-many further steps in S. It is quite obvious that there must be a certain relation between cofinal hypotheses (in a revision sequence of length ORD) and reflexive hypotheses. In fact, we can show that these two concepts are essentially equivalent (provided we restrict our attention to revision sequences of length ORD).[4]

[4] Cf. Fact 8.2.1(iv).

(ii) Cofinal (or equivalently reflexive) hypotheses play the central and most important role in the definition of the semantical systems $\mathbf{S}^*$ and $\mathbf{S}^{\#}$[5]. From an intuitive point of view, if we want to know the semantical properties of a revision process, we need to know the hypotheses that occur over and over again. Precisely this is captured in the idea of a recurring hypothesis.

Later, we will see many examples of recurring hypotheses. Therefore, we mention only two examples.

Example 8.1.3 (i) Assume $D = L^+$ is a given language L extended with a truth predicate $\mathbf{T}$. Assume further that $X = \{T, F\}$. Consider the Liar sentence $\lambda = \neg(\mathbf{T}(\ulcorner\lambda\urcorner))$. In every revision sequence S and every hypothesis h, the Liar sentence λ is unstable. Furthermore, there are recurring hypotheses where λ is true and there are recurring hypotheses where λ is false. For each reflexive hypothesis h λ itself is 2-reflexive. Clearly, h itself is not necessarily 2-reflexive.

(ii) Consider the initial example of Section 8.1 once more again:

$$\forall x \in D : G(x) \iff (x = 0) \vee \exists y(G(y) \wedge x = y') \wedge \neg\exists z(G(z) \wedge z' = 0)$$

Clearly, $G = \mathbb{N}$ is a recurring hypothesis because this hypothesis is stable for every further revision. Are there other recurring hypotheses? The answer is no. Every revision enlarges the extension of G until the fixed point $\mathbb{N}$ is reached, except we start with an extension that contains already $\mathbb{N}$. Then, the revision process stabilizes after the next revision (at the fixed point $\mathbb{N}$).

Now, we can state the central definitions for validity of the two semantical systems $\mathbf{S}^*$ and $\mathbf{S}^{\#}$. (Although we will not study $\mathbf{S}^{\#}$ extensively, we add the definition of this system, too.) In the following definition, a classical model $\mathfrak{M}$ is an ordinary (classical) model relative to a two-valued classical logic and the standard Tarski semantics for a given language L.[6]

Definition 8.1.4 *(i) Let L be a language, $\mathfrak{M}$ be a classical model, D be a (not necessarily circular) definition specifying a predicate P, and ϕ be a sentence of $L^+ = L \cup \{P\}$. Then validity of a sentence ϕ in $\mathfrak{M}$ on D in $\mathbf{S}^*$ is defined as follows:*

$$\mathfrak{M} \models^* \phi \iff \forall h : Ref_D(h) \rightarrow \mathfrak{M} + h \models \phi$$

(ii) Assume the premises of (i). Then: ϕ is valid on D in $\mathbf{S}^*$ is defined as follows:

$$\models^* \phi \iff \forall \mathfrak{M} : \mathfrak{M} \models^* \phi$$

[5] Cf. Definition 8.1.4.

[6] Compare [ChKe73] for further information concerning classical model theory.

(iii) Assume the same premises as in (i). Then: The sentence ϕ is valid in $\mathfrak{M}$ on D in $\mathbf{S}^{\#}$ is defined as follows:

$$\mathfrak{M} \models^{\sharp} \phi \iff \forall h \exists n \in \omega \forall p \geq n : Ref_D(h) \rightarrow \mathfrak{M} + \rho^p(h) \models \phi$$

(iv) Assume the same premises as in (ii). Then, ϕ is valid on D in $\mathbf{S}^{\#}$ is defined as follows:

$$\models^{\sharp} \phi \iff \forall \mathfrak{M} : \mathfrak{M} \models^{\sharp} \phi$$

Remark 8.1.4 We worked loosely with definitions D and revision rules ρ. It is clear from the context what we mean when we write $Ref_D(h)$: h is a hypothesis that is reflexive and furthermore is a hypothesis for ρ, where ρ is induced by definition D. Clearly, every definition D induces a unique revision rule ρ. Hence, when we write D, the revision rule ρ is uniquely determined.

There are other possibilities to define semantical systems for circular definitions. Because semantical validity in $\mathbf{S}^*$ and $\mathbf{S}^{\#}$ is essentially defined via recurring revision sequences and therefore deals with revision sequences of length ORD, it is a relatively obvious idea to define semantical systems via revision sequences that are only of length α (where α is an ordinal). Furthermore, one could restrict the considerations to n-reflexive hypotheses for $n \in \omega$. Roughly speaking, these changes result in the systems $\mathbf{S}_n$. Because the properties of these systems are well-known from [GuBe93, Kr93] and [An94a], we restrict our attention to the stronger systems $\mathbf{S}^*$ and $\mathbf{S}^{\#}$.

In the next section, we will summarize some facts concerning Gupta-Belnap systems. We do not try to give a complete overview of the various systems. We will only mention the standard results which are important for an understanding of the further argumentation and possible applications. Furthermore, we will mention the properties that are crucial for the theory itself.

8.2 Facts and Examples

Some important properties of revision sequences and semantical systems can be found in [GuBe93]. It is clear that we cannot give a complete overview of the theory in this section. We begin with some simple facts concerning revision sequences, recurring hypotheses, and validity in the semantical systems $\mathbf{S}^{\#}$ and $\mathbf{S}^*$. As above we suppress quite often the subscripts referring to a particular revision rule ρ in a given revision sequence.

Fact 8.2.1 *(i) Every revision sequence S of length ORD includes at least one cofinal hypothesis h.*
(ii) Let S be a revision sequence. If $d \in D$ is stably $x \in X$, then for every cofinal hypothesis h in S it holds: $h(d) = x$.
(iii) If a hypothesis h is cofinal in a revision sequence S, then $\rho(h)$ is cofinal in S, too.

(iv) Assume we work in revision sequences that are of length ORD, then a hypothesis h is reflexive if and only if h is cofinal.
(v) If ϕ is valid in $\mathfrak{M}$ in $\mathbf{S}^*$, then ϕ is valid in $\mathfrak{M}$ in $\mathbf{S}^{\#}$. The converse is not generally true.

Proof: (i) The argument for the claim is a simple calculation concerning the involved cardinalities. Consider the set $\{h \mid h \text{ is not cofinal in } S\}$ of all hypotheses that are not cofinal in S. Furthermore, consider the ordinal $\alpha = \max\{\alpha_h \mid h$ does not occur in S after $S_{\alpha_h}\}$. This ordinal α exists because we have at most $|D^X|$ many different hypotheses. Then it holds for all $\beta > \alpha : S_\beta$ is cofinal.

(ii) Obvious from the definitions.

(iii) Because h is cofinal in S, there exists an α, such that $S_\alpha = h$. Then: $S_{\alpha+1} = \rho(h)$. After β many revisions, where β is an appropriate ordinal, we have: $S_{\alpha+\beta} = h$, and therefore $S_{\alpha+\beta+1} = \rho(h)$ again. Conclude: $\rho(h)$ is cofinal in S.

(iv) "$\Rightarrow$" Assume h is α-reflexive in S. By definition it holds: there is a revision sequence S, such that $S_0 = S_\alpha = h$. Now consider a revision sequence S' of length ORD, such that $h = S_0 = S_\alpha = S_{\alpha+\alpha} = S_{\alpha+\alpha+\alpha} = \ldots$. Obviously, h is cofinal by definition.
"$\Leftarrow$" Assume h is cofinal in S. Assume that for $\alpha \in ORD$ it holds: $S_\alpha = h$. Construct a new revision sequence S', such that it holds for all β: $S'_\beta = S_{\alpha+\beta}$. Because h is cofinal, we have: there exists $\gamma \in ORD$, such that it holds: $S_{\alpha+\gamma} = h$. Therefore it holds: $S'_\gamma = h$. But then: $S'_0 = S'_\gamma = h$. Conclude: h is reflexive.

(v) Assume it holds $\mathfrak{M} \models^* \phi$ (relative to a given circular definition D). By definition, ϕ is true in all classical models $\mathfrak{M} + h$ where h is an arbitrary cofinal hypotheses for a revision sequence S. Take $p = 0$ in the definition of validity in $\mathbf{S}^{\#}$. Then it holds: $\mathfrak{M} \models^\sharp \phi$ (relative to D).
It remains to show that the converse is not necessarily true. Consider the following example: Let G be a predicate defined as follows (where our domain is given by the set of natural numbers $\mathbb{N}$):

$$G(x) \Leftrightarrow [G \text{ closed } \wedge \ x \neq x] \ \vee \ [\neg G \text{ closed } \wedge \ \forall y \in \omega(y < x \to G(y))]$$

In this context, the expression 'G is closed' means that the following condition holds:

$$\forall x[\forall y(y < x \to G(y)) \to G(x)]$$

The formula $\neg\forall x(G(x))$ is clearly not valid in $\mathbf{S}^*$ because for the the recurring hypothesis $h = \mathbb{N}$, the formula $\neg\forall x(G(x))$ is not true. But $\neg\forall x(G(x))$ is valid in $\mathbf{S}^{\#}$ because even for $h = \mathbb{N}$, we can take $n = 1$, such that for all $p \geq n$ the formula $\rho^p(h)$ makes $\neg\forall x(G(x))$ true. q.e.d.

The differences between the semantical system $\mathbf{S}^*$ and the semantical system $\mathbf{S}^\#$ are not immediately obvious. The above example in the proof of Fact 8.2.1(v) illustrates that there is a difference between the considered systems. As a matter of fact from a global perspective the differences between $\mathbf{S}^\#$ and $\mathbf{S}^*$ diminish, as we will see later: the complexity of validity in both systems is Π^1_2.

The importance of the following theorem is based on a variety of consequences for different applications of revision theories. This theorem is usually called McGee's Theorem. It was first proven in [Mc91]. Our presentation is based on [GuBe93]. McGee's theorem specifies an upper bound of the length of a circle in the chain of hypotheses, i.e. an upper bound for a reflexive hypothesis. This upper bound of a revision rule $\rho : X^D \longrightarrow X^D$ is essentially restricted by the domain D and the range of the revision rule ρ.

Theorem 8.2.2 *Assume h is a reflexive hypothesis for a given revision rule $\rho : X^D \longrightarrow X^D$. Then h is α-reflexive for an ordinal α such that $|\alpha| \leq \max(|D|, |X|, \aleph_0)$.*

Proof: We only sketch the proof of this theorem. Suppose h is β-reflexive for a revision sequence S and an ordinal β. We use a Löwenheim-Skolem argument to construct a revision sequence S', such that h is α-reflexive in S' with $|\alpha| \leq \max(|D|, |X|, \aleph_0)$. Assume L is a first-order language with constants ORD, Obj, Val, $Less$, and R. Let $\mathfrak{M} = \langle D, I\rangle$ be a classical ground model with $D = X \cup D \cup \{\gamma \mid \gamma \leq \beta + \omega\}$. The interpretation of the constants given above are defined as follows:

$I(ORD)(d) = t \;\Leftrightarrow\; d$ is an ordinal
$I(Obj)(d) = t \;\Leftrightarrow\; d \in D$
$I(Val)(d) = t \;\Leftrightarrow\; d \in X$
$I(Less)(d, d') = t \;\Leftrightarrow\; d$ and d' are ordinals and furthermore $d < d'$
$I(R)(\alpha, d, x) = t \;\Leftrightarrow\; S_\alpha(d) = x$

Applying the Löwenheim-Skolem theorem (downwards), there is a model $\mathfrak{M}' = \langle D', I'\rangle$ of cardinality κ, such that $\mathfrak{M}'$ is an elementary submodel of $\mathfrak{M}$. It is clear that D' includes D, X, the set $\{0, \beta\}$, and a subset Γ of the ordinals. Γ is in general not an initial segment of the ordinals. Hence, we define an isomorphism f mapping Γ into an initial segment of the ordinals according to the following condition:

$$f(\gamma) = \bigcup\{f(\delta) + 1 \mid \delta \in \gamma \cap D'\}$$

It is easy to check that f is an isomorphism. Define the new revision sequence S' as follows (here we use essentially the interpretation of the constant R given above):

$$(\forall \gamma \leq \alpha + \omega)(\forall d \in D)(\forall x \in X) : S'_\gamma(d) = x \;\Leftrightarrow\; I'(R)(f^{-1}, d, x) = t$$

Let be $\alpha = f(\beta)$. Then: $|\alpha| \leq \max(|D|, |X|, \aleph_0)$. It remains to show that S' is a revision sequence. The first condition for revision sequences concerning successor stages is obviously fulfilled by the fact that $\mathfrak{M}'$ is an elementary submodel of $\mathfrak{M}$. Concerning the limit stage we reason as follows: if γ is a limit ordinal and d is stably x in $S \upharpoonright \gamma$, then $d = f^{-1}(g)$ is a limit ordinal, too. Additionally, the following relation holds in $\mathfrak{M}'$:

$$\exists y(ORD(y) \ \wedge \ Less(y, x_1) \wedge (\forall z(ORD(z) \ \wedge \ Less(z, x_1) \ \rightarrow \ R(z, x_2, x_3))))$$

Then, the above relation holds in $\mathfrak{M}$ as well (because $\mathfrak{M}'$ is an elementary submodel of $\mathfrak{M}$). Finally, because S is a revision sequence, we have: $S_\delta(d) = x$. Conclude: $S'_\gamma(d) = x$. q.e.d.

Remark 8.2.1 Although McGee's theorem is essentially based on a classical Löwenheim-Skolem argument and therefore straightforward to prove, this claim is very important for the further development of revision theories. It shows that given an α-reflexive hypothesis h in a revision sequence S (for a revision rule ρ), we can find a revision sequence S' for revision rule ρ where h is recurring after minimal many steps. That simplifies many revision theoretic applications. One application is the problematic feature of revision theories to deal with sequences of hypotheses that have length ORD. Then, a single revision sequence S is a proper class and no longer a set. With McGee' Theorem, it is possible to show that in fact we only need sets to code the theory and get the same expressive power.[7]

To summarize important results of revision theories, we shall mention a further semantical system, namely the inductive system $\mathbf{S}_i$, specifying the concept of inductive validity. This notion is not only important for theoretical reasons, but it is the key to prove most important theorems about the Gupta-Belnap systems $\mathbf{S}^*$ and $\mathbf{S}^\#$, below. Using the inductive system $\mathbf{S}_i$, one can show that the systems $\mathbf{S}^*$ and $\mathbf{S}^\#$ are not recursively axiomatizable. The following definition makes the semantical system $\mathbf{S}_i$ precise.

Definition 8.2.3 *Assume L is a given first-order language and L^+ is the expansion of L with a finite list of inductively defined predicates $\{P_1, P_2, ..., P_n\}$. We define: $\phi \in L^+$ is inductively valid on inductive definitions $\{D_1, D_,...D_n\}$ in $\mathfrak{M}$ (we will use the notation $\mathfrak{M} \models_i \phi$) if and only if ϕ is valid in $\mathfrak{M} + \{P_1, P_2, ..., P_n\}$ where every P_i is interpreted as the extension of the minimal fixed point of the inductive definition D_i. If ϕ is inductively valid for every model $\mathfrak{M}$, then we call ϕ inductively valid. We call the resulting semantical system $\mathbf{S}_i$.*

[7] Cf. Section 10.2 for further remarks concerning this point.

In the remaining part of this section, we examine some examples of reflexive (recurring) hypotheses and their properties. These examples should clarify the given definitions.

Example 8.2.2 (i) Assume we are working in a language with the signature $\{0,', <\}$ where 0 is a constant, $'$ is a set theoretical successor function, and $<$ is a strict linear well-order with 0 as the smallest element. Consider the following circular definition:

$$G(x) \iff [x = 0 \lor \exists y(G(y) \ \land \ y' = x)]$$

It is easy to see that the reflexive hypotheses are initial segments of the ordinals closed under the successor operation, such that limit ordinals are excluded. If at a certain step the extension of G is such an initial segment of the ordinals which is not closed, the next revision step adds another ordinal to the extension of G. Notice that we model an inductive process with the above definition of G. In the case of a given domain D, we have the following situation: If $D = \mathbb{N}$, then there is only one reflexive hypothesis, namely $h = \mathbb{N}$. If we work in some initial segment of the ordinals, say up to λ (where λ is a limit ordinal), then every initial segment up to a limit ordinal $\alpha < \lambda$ gives us a reflexive hypothesis, provided all limit ordinals β with $\beta < \alpha$ are excluded. Notice that the revision becomes stable after at most α many steps where $\alpha \geq |D|$.

(ii) Now assume we work in a given domain $D = \mathbb{Z} = \{\cdots - 2, -1, 0, 1, 2, \ldots\}$ and a signature $\{0,', <\}$.[8] Again, we want to know the reflexive hypotheses of the circular definition from (i):

$$G(x) \iff [x = 0 \lor \exists y(G(y) \ \land \ y' = x)]$$

Obviously, the natural numbers $\mathbb{N}$ and the integers $\mathbb{Z}$ (i.e. the whole domain) are reflexive hypotheses because both remain stable during arbitrarily many further revision steps. Furthermore, there exist infinitely many more reflexive hypotheses, namely all subsets $\mathbb{N} \cup X$ where X is an infinite subset of $\mathbb{Z} - \mathbb{N}$, such that $X^C - \mathbb{N}$ is infinite. This is clear for cases like the following hypothesis h (where we designate numbers that are in the extension of G with italic numbers):

$$h \ = \ \langle \ldots \textit{-6}, -5, \textit{-4}, -3, \textit{-2}, -1, \textit{0}, \textit{1}, \textit{2}, \textit{3}, \ldots \rangle$$

The next revision results in the following extension for G:

$$\rho(h) = \{\cdots - 5, -3, -1, 0, 1, 2, \ldots\}$$

A further application of the revision rule yields again the set

$$\rho(\rho(h)) = \{\cdots - 4, -2, 0, 1, 2, \ldots\}$$

We should mention that the following h' is also a reflexive hypothesis, although there is no finite number $n \in \mathbb{N}$, such that $\rho^n(h') = h'$. That means that

[8] Expand the extension of the successor operation $'$ to negative integers. This can be done straightforwardly.

the following hypothesis h' is not n-reflexive for a natural number $n \in \mathbb{N}$, but α-reflexive for an appropriate ordinal α.

$$h' \;=\; \langle \dots \textit{-10}, -9, -8, -7, \textit{-6}, -5, -4, \textit{-3}, -2, \textit{-1}, \textit{0}, \textit{1}, \textit{2}, \textit{3}, \dots \rangle$$

Although we cannot find a finite number of revisions to get hypothesis h', we can find a ω-reflexive revision sequence S such that h' is one of the reflexive hypotheses. This is possible by choosing at every limit stage of the revision process for the unstable elements (i.e. for the negative integers) the initial hypothesis. Notice: given $\mathbb{Z}$ as the domain of our considerations, there exist $\aleph_1$ many reflexive hypotheses.

(iii) Assume we are working in the signature $\{0,', <\}$. We want to define the natural numbers using circular definitions in the sense that the circular definition for a predicate H has only a unique reflexive hypothesis h that is coextensional to the natural numbers.[9] Consider the following two definitions:

$$\begin{aligned} G(x) \iff & \forall x : (G(x) \to [x = 0 \lor \exists y(G(y) \land y' = x)])\land \\ & \exists x \neg([x = 0 \lor \exists y(G(y) \land y' = x)] \to G(x))\land \\ & [x = 0 \lor \exists y(G(y) \land y' = x)] \end{aligned}$$

$$\begin{aligned} H(x) \iff & [\forall x(G(x) \to [x = 0 \lor \exists y(G(y) \land y' = x)])\land \\ & \forall x([x = 0 \lor \exists y(G(y) \land y' = x)] \to G(x)) \land G(x)]\lor \\ & [\neg(\forall x(G(x) \to [x = 0 \lor \exists y(G(y) \land y' = x)])\land \\ & \forall x([x = 0 \lor \exists y(G(y) \land y' = x)] \to G(x))) \land H(x)] \end{aligned}$$

Checking the possible initial hypotheses for G, we find the following properties. If the initial hypothesis for G is the empty set, then the revision process is mirroring induction steps. After precisely ω-many steps we have an isomorphic copy of the natural numbers as the extension of G. The revision revises the extension of G to the empty set again and the whole process starts again. Starting with a hypothesis that is not an initial segment of the ordinals, the next revision determines the extension of G to the empty set and the revision starts with the above scenario. The definition for G is an interesting example of a definition that never settles at a fixed point. The behavior of G is in a sense similar to Liar-like sentences, except that the revisions take not two steps to reach the original hypothesis but ω-many steps. The predicate H shows the following behavior. It remains fixed, except the extension of G is an isomorphic copy of the natural numbers. Then, H flips to the extension of G. The following revisions do not change the extension of H: the natural numbers are a fixed point for the circular definition for H. It is clear that $\mathbb{N}$ is the only reflexive hypothesis for H, whereas every initial segment of ω is reflexive for G. This example is a special case for a very general modeling of

[9] That corresponds not precisely to the notion *strong definability* in [GuBe93]. In [GuBe93], only near stability is required. By interchanging $\mathbf{S}^{\#}$ and $\mathbf{S}^{*}$ both notions are equivalent.

inductive definitions via revision sequences. It turns out that revision theories can represent every inductive definition. We will consider the special case of Example 8.2.2(iii) in different contexts below.

(iv) Assume we are working in the language $\langle 0, <, ', +, \cdot \rangle$ and our domain is the set of the natural numbers ω. Consider the following definition of a subset of ω:

$$G(x) \iff \exists y\,(x = 2y)$$

This definition gives us obviously the even numbers, because x can only be in the extension of G in the case that x is equal to two times y for a y to be a natural number. Now, we consider a different definition:

$$\begin{aligned} G'(x) \iff & [(G'(x) \wedge (\exists y\,(x = 2y))) \rightarrow (\neg G'(x))] \\ & \wedge\ [(\neg G'(x) \wedge (\exists y\,(x = 2y))) \rightarrow G'(x)] \end{aligned}$$

What are the reflexive hypotheses of the last example? The odd numbers are in the extension of G' in every revision stage because odd numbers make the antecedents of the two conditional formulas false. If x is an even number, and in the extension of G', then in the next revision x is not in the extension of G' and vice versa. Therefore, we have as reflexive hypotheses the set of the natural numbers $\mathbb{N}$ in the case that the even numbers are in the extension of G' and additionally the set $\{n \in \omega \mid \exists m \in \omega : n = 2m + 1\}$ in the case the even natural numbers are not in the extension of G'. Clearly, the only stable elements are the odd numbers, whereas the even numbers are 2-reflexive.

The first two examples are taken from [GuBe93], whereas the third example is a special case of a general connection between inductive definitions and circular definitions that goes back to [Kr93]. We finish our examples here, although many interesting properties could be additionally exemplified. For a large number of non-trivial examples the reader is referred to [GuBe93].

In Definition 8.1.4, we specified the semantical systems $\mathbf{S}^*$ and $\mathbf{S}^\#$. We add some examples for validity in $\mathbf{S}^*$. As in the case of recurring hypotheses above, we will see further examples later, therefore we only try to give a flavor of the properties of these systems.

Example 8.2.3 (i) Consider Example 8.2.2(i). The sentence ϕ specified as

$$\phi \ = \ (\forall x : x \in G \leftrightarrow \exists y \in D : y' = x)$$

is true for every reflexive hypothesis. Therefore, we have: $\models^* \phi$. ϕ is valid because there is no reflexive hypothesis that contains a limit ordinal.

(ii) Consider Example 8.2.2(ii). A sentence of the form $\phi = \neg\forall x\,(G(x))$ is not valid, because for the reflexive hypothesis $h = \mathbb{Z}$ it holds $\mathfrak{M} + h \not\models^* \phi$. Notice that from $\not\models^* \phi$ it does not follow that $\models^* \neg\phi$. The latter expression would only be true, if $\mathbb{Z}$ was the unique recurring hypothesis.

(iii) Consider again Example 8.2.2(ii). If we take ϕ to be the sentence

$$\phi \;=\; [\forall m \in \mathbb{Z}\, \exists x(G(x) \wedge (\exists n \in \mathbb{N} : x^n = m))]$$

is not valid, i.e. $\not\models^* \phi$. This holds because for a negative number m and the reflexive hypothesis $h = \mathbb{N}$, ϕ cannot be true.

(iv) All classical tautologies remain true in $\mathbf{S}^*$. That holds because all classical statements are interpreted relative to the classical ground model $\mathfrak{M}$. Only circular definitions are revised via the properties of revision sequences based on the induced revision rule.

(v) Consider the circular definition of G' in Example 8.2.2(iv). The sentence $\phi \;=\; (\forall x : G'(x) \;\leftrightarrow\; x = 2m + 1)$ is not valid in $\mathbf{S}^*$. This shows that the definition of G' does not strongly define the odd numbers. For example, the following holds:

$$\models^* \forall x : x = 2m + 1 \;\rightarrow\; G'(x)$$

In other words, the property to be an odd number is a sufficient condition to be in the extension of G'. On the other hand, it holds:

$$\not\models^* \exists x : G'(x) \wedge x \neq 2m$$

Referring to its behavior one can call such a definition a weak definition of the odd numbers.

Our further interest in this chapter is devoted to results concerning the complexity theory of the semantical systems $\mathbf{S}^*$ and $\mathbf{S}^{\#}$. The focus of these considerations will be to determine the complexity of the collection of theorems of the system $\mathbf{S}^*$.

8.3 Some Results concerning Revision Theories

We will state a theorem due to Kremer[10] which is one of the main results in the theory of Gupta-Belnap systems. In this section, we will work in arithmetic. Let L be the language $\{0, <, +, \cdot, '\}$, assume that PA^- is a shorthand for the axioms of Peano arithmetic without the induction axiom, and let SLO be a set of axioms which guarantees that $<$ is a strict linear order with a smallest element 0. The successor function $'$ is defined as usual. Assume $\mathbb{N}$ denotes standard arithmetic. Notice that we used $\mathbb{N}$ to denote the set of natural numbers in the above sections. We mention a theorem by Kremer that gives us a first approximation of the complexity of the inductive system $\mathbf{S}_i$. (Inductive systems

[10] Cf. [Kr93].

are defined in Definition 8.2.3.) This result is used later for a calculation of the complexity of revision theories, more precisely of the complexity of the semantical system $\mathbf{S}^*$.

Lemma 8.3.1 *Assume $A \subseteq \omega$ is a Π^1_2 subset of the natural numbers. Then: A is recursively embeddable into $\{\phi \mid \phi$ is valid on D in $S_i\}$ for an appropriate finite list D of inductive definitions.*

Proof: Assume that $A \subseteq \omega$ is a Π^1_2 subset of the natural numbers. Then, the following equivalences hold by logic:

$$\begin{array}{rl} & x \in A \;\Leftrightarrow\; \forall X \exists Y : R(X, Y, x) \\ \text{iff} & x \notin A \;\Leftrightarrow\; \neg\forall X \exists Y : R(X, Y, x) \\ \text{iff} & x \notin A \;\Leftrightarrow\; \exists X \forall Y : \neg R(X, Y, x) \end{array}$$

In the above formula, R is considered an arithmetical relation. Because A is a Π^1_2 relation, the complement of A is Σ^1_2. Therefore, the set A^C can be specified as follows.

$$(1)\;\; A^C = \{n \mid n \in \omega \;\wedge (\exists X \subseteq \omega \times \wp(\omega) \forall Y \subseteq \omega : \langle n, Y\rangle \notin X)\}$$

Notice that from (1) it follows that X is a Π^1_1 predicate. Assume Q is a predicate that is inductively defined and assume further that the extension of the minimal fixed point of the inductive definition of Q gives us the interpretation of the variable X. Assume further that H is defined as follows:

$$H(x) \;\Leftrightarrow\; x = 0 \vee \exists y (H(y) \;\wedge\; y' = x)$$

It is clear that the circular definition of H precisely defines the natural numbers. It should be noticed that the reason for the usage of H is to find a way to eliminate the unrestricted quantification over models in the proof below. With these specifications of Q and H, it is clear that the following equivalence holds:

$$n \notin A \;\Leftrightarrow\; \models_i PA^- \wedge SLO \wedge \forall x : (Hx \rightarrow \neg Q(n))$$

Because Q specifies X and H defines the natural numbers, the following equivalence holds:

$$(2)\;\; (\forall n \in \omega)(\forall Y \subseteq \omega) : [\langle n, y\rangle \notin X] \;\Leftrightarrow\; [\mathbb{N} + Y \models_i \neg Q(n)]$$

In the following, we show that inductive validity in $\mathbb{N}$ is coextensional to validity in $PA^- \wedge SLO \wedge H$. First, we consider the left to right direction.

$$(3)\;\; [\forall Y \subseteq \omega : \mathbb{N} + Y \models_i \phi] \;\Rightarrow\; \models_i PA^- \wedge SLO \wedge \forall x (Hx \rightarrow \phi)$$

Assume that it holds: $\not\models_i PA^- \wedge SLO \wedge \forall x (Hx \rightarrow \phi)$. Then, there is a model $\mathfrak{M} = \langle D, I\rangle$ and a subset Y of ω, such that $\mathfrak{M}+Y \not\models_i PA^- \wedge SLO \wedge \forall x (Hx \rightarrow \phi)$. This is equivalent to the following statement:

$$\begin{aligned}\mathfrak{M} + Y + H + Q &\models PA^- \wedge SLO \wedge \forall x Hx \qquad \text{and}\\ \mathfrak{M} + Y + H + Q &\not\models \phi\end{aligned}$$

Because Hx holds for the whole domain D of $\mathfrak{M}$ (unrestricted quantification), D must have the same extension as the minimal fixed point of the inductive definition of H, i.e. D must be isomorphic to the natural numbers ω. Hence, there is an isomorphism α mapping the domain of the model into the natural numbers ω. Then, it can be established:

$$\mathbb{N} + \alpha(Y) + \alpha(H) + \{\alpha(d) \mid d \in D \wedge Q(d)\} \;\cong\; \mathbb{N} + Y + H + Q$$

Using α it holds:

$$\exists \alpha(Y) : \alpha(Y) \subseteq \omega \;\wedge\; \mathbb{N} + \alpha(Y) \not\models_i \phi$$

Using the relations specified in (1), (2), and (3) the following chain of implications hold:

$$\begin{aligned} & \forall n \in \omega : (n \notin A)\\ \Rightarrow\quad & \forall Y \subseteq \omega : \langle n, Y\rangle \notin X\\ \Rightarrow\quad & \forall Y \subseteq \omega : \mathbb{N} + Y \models_i \neg Q(n)\\ \Rightarrow\quad & \models_i PA^- \wedge SLO \wedge \forall x(Hx \rightarrow \neg Q(n))\end{aligned}$$

This proves one direction of the lemma.

To complete the proof we have to show that the reverse direction of expression (3) holds. We want to prove that the following relation holds:

$$(4)\quad \models_i PA^- \wedge\; SLO \wedge \forall x(Hx \rightarrow \phi) \;\Rightarrow\; [\forall Y \subseteq \omega : \mathbb{N} + Y \models_i \phi]$$

As above we use a proof by contraposition. Assume towards a contradiction that there is a subset $Y \subseteq \omega$, such that $\mathbb{N} + Y \not\models_i \phi$. Then it holds by definition: $\mathbb{N} + Y + Q + H \not\models \phi$. On the other hand, we have: $\mathbb{N} + Y + Q + H \models PA^- \wedge SLO \wedge \forall x Hx$ (because $\mathbb{N}$ is the standard model of arithmetic). Ordinary logic yields the relation:

$$\mathbb{N} + Y + Q + H \;\not\models\; PA^- \wedge\; SLO \wedge \forall x(Hx \rightarrow \phi)$$

This means:

$$\mathbb{N} + Y \;\not\models_i\; PA^- \wedge SLO \wedge \forall x(Hx \rightarrow \phi)$$

Then we have

$$\not\models_i\; PA^- \wedge SLO \wedge \forall x(Hx \rightarrow \phi)$$

Now, we can apply the following reasoning, using the relations (1), (2), and (4) similarly as above:

$$\begin{array}{rl} & \models_i PA^- \wedge SLO \wedge \forall x(Hx \rightarrow \neg Q(n) \\ \Rightarrow & (\forall Y \subseteq \omega : \mathbb{N} + Y \models_i \neg Q(n)) \\ \Rightarrow & \forall Y \subseteq \omega : \langle n, Y \rangle \notin X \\ \Rightarrow & \forall n \in \omega : (n \notin A) \end{array}$$

We have proven that an arbitrary Π^1_2 subset of ω can be recursively embedded into the set $\{\phi \mid \phi \text{ is valid on } D \text{ in } \mathbf{S}_i\}$. This completes the proof of the lemma.

q.e.d.

Lemma 8.3.1 tells us a fact about the complexity of systems that are enriched with finitely many additional inductive definitions (more precisely with two additional inductive definitions). The idea was to use an inductive definition (namely the definition of H) to model the natural numbers in order to avoid the unrestricted quantification over arbitrary models, and to use another inductive definition to model a Π^1_1 predicate. Then, the whole consideration reduces to an application of ordinary logical techniques.

Although we know by Lemma 8.3.1 something about the complexity of inductive systems $\mathbf{S}_i$, we do not know anything about the complexity of the semantical systems $\mathbf{S}^*$ or $\mathbf{S}^\#$ that are primarily in the focus of our consideration. We shall show that every inductively defined predicate P can also be defined in $\mathbf{S}^*$ and $\mathbf{S}^\#$. Then, it is clear that the complexity of revision theoretic validities (w.r.t. the semantical systems $\mathbf{S}^*$ and $\mathbf{S}^\#$) must be at least the complexity of the inductively enriched first-order systems of Lemma 8.3.1, i.e. they must have at least complexity Π^1_2. First, we show that every inductive definition can be modeled by a circular definition, such that this definition has only one recurring hypothesis. The proof goes back to unpublished ideas by Anil Gupta.

Lemma 8.3.2 *Let D be an inductive definition for a predicate P, i.e. let Γ be a monotone operator, such that the minimal fixed point of Γ determines a unique extension of the inductively defined predicate P. Then it holds: there is a circular definition D' for a predicate H, such that $x \in P$ if and only if x is in the extension of a unique reflexive hypothesis h for D' (relative to a revision rule ρ induced by definition D').*

Proof: We show that the application of a very general (circular) definition in $\mathbf{S}^*$, guarantees to define every inductive predicate.[11] Assume X is an inductively defined subset of ω, such that $A(x, G)$ denotes the definiens of the inductive definition of X. We want to find circular definitions, such that a certain predicate H has the same extension as X. In order to do this, consider the following definitions for the predicates G and H:

[11] The following pair of circular definitions is called 'Gupta translation' in [Kr93].

$$G(x) \Leftrightarrow \forall x(G(x) \to A(x,G)) \wedge \exists x \neg (A(x,G) \to G(x)) \wedge A(x,G)$$

$$\begin{aligned} H(x) \Leftrightarrow & \quad [\forall x(G(x) \to A(x,G)) \wedge \forall x(A(x,G) \to G(x)) \wedge G(x)] \\ & \vee \ [\neg(\forall x(G(x) \to A(x,G)) \wedge \forall x(A(x,G) \to G(x))) \wedge H(x)] \end{aligned}$$

First, we state the properties of G. If our initial hypothesis is the empty set, then after the first revision the base case of the inductive definition is in the extension of G. Notice that $\forall x(G(x) \to A(x,G))$ holds vacuously and $\exists x \neg (A(x,G) \to G(x))$ is true for the base case. After that, the revision process extents the extension of G step by step like an ordinary inductive definition through the natural numbers. In particular, it holds:

$$\exists \lambda \in ORD \forall x: \ x \in X \ \Leftrightarrow \ x \in \rho^{\lambda}(\emptyset)$$

Hence, we reached a fixed point of the inductive process. The $\lambda + 1^{st}$ revision forces the extension of G to be the empty set, because $\forall x(A(x,G) \to G(x))$ holds. Therefore, for every x the conjunction of the definiens is false. The whole process starts again.
In the case that the initial hypothesis is neither the empty set nor an initial segment of the natural numbers, the first revision step forces the extension of G to be $\emptyset$ (because of the expression $\forall x(G(x) \to A(x,G))$). Therefore, after at most $\lambda + 1$ many steps the revision process will reach a fixed point (by the considerations above). It is important to notice that the revision rule ρ is monotone increasing after one revision. After at most $\lambda + 1$ many revisions, ρ is decreasing (mapping X to $\emptyset$) and after that ρ is monotone increasing again. In a certain sense, the circular definition is 'nearly' an inductive definition.
The behavior of H can be summarized as follows. In the case that it holds $\forall x(G(x) \leftrightarrow A(x,G))$, the extension of H is the extension of the minimal fixed point $\rho^{\lambda}(\emptyset)$ of the inductive process. In all other cases, the extension of H remains fixed whatever the extension before the revision was. Therefore, after at most $\lambda + 1$ many revisions, the extension of H reaches a fixed point and remains stable for further arbitrary revisions . Conclude: there is only one recurring hypothesis for H, namely the extension of the inductively defined predicate P. q.e.d.

As a corollary of the above lemma, it is easy to see that every coinductive definition can be mirrored by a revision theoretic definition. This is a consequence of the fact that for every inductively defined predicate, we can also define the negation of this predicate. Because the negation of every Π^1_1 relation is a Σ^1_1 relation the claim follows immediately. The following corollary specifies the situation in the coinductive case.

Corollary 8.3.3 *Let D be a coinductive definition (Σ^1_1 definition) inducing a maximal fixed point of a monotone operator Γ. Let P be the predicate uniquely* *defined by this fixed point. Then it holds: there is a revision theoretic definition D' defining a predicate H, such that $x \in P$ if and only if x is in the extension of a unique reflexive hypothesis h for H (relative to a given revision rule ρ induced by definition D').*

Proof: Assume X is a coinductive subset of ω. Then, X is a Σ^1_1 subset of ω. The complement of X is a Π^1_1 set X^C. We know by Lemma 8.3.2 that X^C is coextensional to a predicate H that can be defined by a revision theoretic definition D defining a predicate H (using another revision theoretically definable relation G). In other words: the unique reflexive hypothesis h of D gives us the extension of X^C. Define a new predicate H' by:

$$x \in H' \;\Leftrightarrow\; x \notin H$$

Then, the following equivalences are true (by ordinary logical reasoning):

$$x \in X \;\Leftrightarrow\; x \notin X^C \;\Leftrightarrow\; \neg H(x) \Leftrightarrow H'(x)$$

Because H has a unique reflexive hypothesis h, H' also has a unique reflexive hypothesis h'. Conclude: h' is coextensional with X. q.e.d.

As an immediate consequence it follows that the semantical systems $\mathbf{S}^*$ and $\mathbf{S}^\#$ are at least as complex as the inductive system $\mathbf{S}_i$. The next corollary states that result.

Proposition 8.3.4 *The complexity of* $\mathbf{S}^*$ *and* $\mathbf{S}^\#$ *is at least* Π^1_2.

Proof: This is a trivial consequence of Lemma 8.3.1 and Lemma 8.3.2. By Lemma 8.3.2 we have: Every inductive definition can be mirrored in $\mathbf{S}^*$ (and $\mathbf{S}^\#$, respectively) by a unique reflexive hypothesis h. Therefore, we can define a finite list of inductive definitions using circular definitions. The framework concerning inductive systems is purely classical. Therefore, we can model validity of the inductive system $\mathbf{S}_i$ in $\mathbf{S}^*$ as well as in $\mathbf{S}^\#$. Because $\mathbf{S}_i$ is at least of complexity Π^1_2, we can conclude that $\mathbf{S}^*$ as well as $\mathbf{S}^\#$ is at least of complexity Π^1_2, too. q.e.d.

We add some remarks concerning some consequences of the above proposition. Although the determination of an upper bound of the complexity class of validity in $\mathbf{S}^*$ seems to be simply a result of mathematical interest, there are a lot of consequences for applications as well.

Remark 8.3.1 (i) From Proposition 8.3.4, it follows immediately that the semantical systems $\mathbf{S}^*$ and $\mathbf{S}^\#$ (as well as the semantical system $\mathbf{S}_i$) are not recursively axiomatizable. This results from the fact that all these systems are more complex than standard arithmetic (which is itself not recursively axiomatizable). As a further consequence it follows that there is no proof-theoretical calculus for these systems.

(ii) At this point, it is important to notice that circular definitions, although they are first-order in nature (i.e. the quantification over variables is consequently restricted to individual variables, hence no quantification over relations or functions is allowed in the definiens of a circular definition) can define second-order concepts, provided they are evaluated in the systems $\mathbf{S}^*$ or $\mathbf{S}^{\#}$. That means intuitively: because of the revision process through all ordinals, the expressive power of these systems increases significantly. Although we will not go into details concerning this topic, an interesting question is which complexity a system of circular definitions would have if we allowed second-order quantifications in the definiens.

(iii) Proposition 8.3.4 gives us a lower bound of the complexity of $\mathbf{S}^*$ and $\mathbf{S}^{\#}$. We shall prove below that this lower bound is in fact optimal, i.e. that the complexity of the systems $\mathbf{S}^*$ and $\mathbf{S}^{\#}$ is Π^1_2-hard. The first (correct) proof of this fact was formulated in [An98].

(iv) Lemma 8.3.2 and Corollary 8.3.3 imply that inductive and coinductive definitions can be strongly defined in $\mathbf{S}^*$ and $\mathbf{S}^{\#}$. In this context, we call a set X strongly definable if and only if the unique reflexive hypothesis h of a circular definition D gives us precisely the extension of X. Notice: This usage of the concept of 'strong definability' differs slightly from the usage in Gupta-Belnap's monograph [GuBe93].

To prove the reverse claim of Proposition 8.3.4 - namely that validity in $\mathbf{S}^*$ is precisely a Π^1_2 relation - we are faced with a problem. Validity in $\mathbf{S}^*$ is defined using recurring (or reflexive) hypotheses and arbitrary models. To decide whether a hypothesis h is reflexive or not, we need to consider sequences of hypotheses that can have uncountable length. In particular, if ranging through all ordinals we are no longer on a set theoretical basis, because each revision sequence is no longer a set. Additionally, the models we have to examine are not necessarily countable, but can have arbitrary cardinality. Therefore, there is no hope to find a characterization of validity of $\mathbf{S}^*$ in the analytic hierarchy by simply formalizing Definition 8.1.4(i)/(ii). Hence, the idea of a characterization is to apply a Löwenheim-Skolem-style argument to find a countable representation of 'validity in $\mathbf{S}^*$' (as well as in $\mathbf{S}^{\#}$). The next lemma claims that validity in $\mathbf{S}^*$ is essentially reducible to the countable case, where the recurring hypotheses are nested in countable revision sequences and the quantification over models can be restricted to countable models. This result was first formulated in [An94a] with a crucial fault in the proof. The proof of this theorem below was first sketched in the unpublished paper [An98].

Lemma 8.3.5 *A sentence ϕ is valid in $\mathbf{S}^*$ if and only if for every hypothesis h which is recurring in a countable revision sequence S and for all countable classical models $\mathfrak{M}$ it holds: $\mathfrak{M} + h \models \phi$.*

Proof: "$\Rightarrow$" Assume ϕ is not valid in $\mathfrak{M} + h$ where $\mathfrak{M}$ is countable and h is a recurring hypothesis in a countable revision sequence S. We have to show that ϕ is not true in $\mathbf{S}^*$. Assume that the premise is true, then the revision sequence S has a certain length, say length α (for a a countable ordinal α). Because every ordinal can be written in the form $\lambda = (\beta \cdot \gamma) + \delta$, we consider the sequence $T_{(\alpha\cdot\gamma)+\delta}$ for $\delta < \alpha$. Then we can associate the following hypotheses of the two revision sequences:

$$\begin{array}{cccccccc}
S_0 & S_1 & \ldots & S_\omega & S_{\omega+1} & \ldots & S_\alpha & S_{\alpha+1} \ \ldots \\
\vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\
T_{(\alpha\cdot 0)+0} & T_{(\alpha\cdot 0)+1} & \ldots & T_{(\alpha\cdot 0)+\omega} & T_{(\alpha\cdot 0)+\omega+1} & \ldots & T_{(\alpha\cdot 1)+0} & T_{(\alpha\cdot 1)+1} \ldots
\end{array}$$

In the revision sequence $T_{(\alpha\cdot\gamma)+\delta}$, the ordinal γ is intended to range through the class of all ordinals. In particular, the resulting revision sequence T has length ORD. The idea is that in the newly defined revision sequence T, the original sequence S is copied over and over again. Obviously, h is reflexive in $T_{(\alpha\cdot\gamma)+\delta}$ because h is reflexive in S. Hence, ϕ is not valid in $\mathbf{S}^*$ because ϕ is not true in $\mathfrak{M} + h$. This is what we wanted to prove.

"$\Leftarrow$": Assume ϕ is false in $\mathfrak{M} + h$ where h is a reflexive hypothesis in a revision sequence based on a given ground model $\mathfrak{M}$. This means that ϕ is false in $\mathbf{S}^*$. We have to show that there is a countable model $\mathfrak{M}'$ and a countable revision sequence S' such that:

(a) h is a recurring hypothesis in S'
(b) ϕ is false in $\mathfrak{M}' + h$.

Now, assume the above premise. According to Theorem 8.2.2 (McGee's theorem) it holds: if h is α-reflexive, then h is also β-reflexive for an ordinal β that is bound by the cardinality of the model $\mathfrak{M}$ or in the case that $\mathfrak{M}$ is finite by $\aleph_0$. Hence, we can assume that S is bound by the cardinality of $\mathfrak{M}$ (which is the only interesting case), say by the ordinal α.
Consider an infinite sequence of models $\langle \mathfrak{M} + S_\beta \rangle$ for $\beta < \alpha$. Because the sequence $\langle \mathfrak{M} + S_\beta \rangle_{\beta<\alpha}$ is bound by the ordinal α and $\mathfrak{M}$ is a model (therefore a set), the sequence $\langle \mathfrak{M} + S_\beta \rangle_{\beta<\alpha}$ is a set in a standard model $\mathcal{M}$ of ZFC, too. Now, we can apply the standard Löwenheim-Skolem theorem to get a countable elementary submodel $\mathcal{M}'$ of $\mathcal{M}$. Because both models of set theory are elementary equivalent (i.e. both make the same sentences true), the following relation holds:

$$\begin{aligned}
\exists \mathcal{M}' \exists \alpha_0 \exists f : \quad & [f : \alpha_0 \longrightarrow \wp(|\mathcal{M}'|)] \\
\wedge\ & [\forall \mathcal{M} : (\mathcal{M} = \mathcal{M}'' + f(\beta+1)) \ \rightarrow\ \mathcal{M} = \delta(\mathcal{M}' + f(\beta))] \\
\wedge\ & [\forall \lambda < \alpha_0 : \lambda \text{ is a limit ordinal} \\
& \ \text{then } f(\lambda) \text{ coheres with } f(\beta) \text{ for all } \beta < \alpha] \\
\wedge\ & [\exists \beta : \beta < \alpha_0 \ \wedge\ \beta > 0 \ \wedge\ f(\beta) = f(0)] \\
\wedge\ & [\mathcal{M}' + f(0) \models \neg\phi]
\end{aligned}$$

It remains to show that S' is in fact a revision sequence. Because the elementary submodel construction preserves the coherence condition, this is obviously satisfied. q.e.d.

Now, we are in the position to show that validity in $\mathbf{S}^*$ has exactly complexity Π^1_2. The idea is quite straightforward: we need to find a formula for the statement 'ϕ is valid in $\mathbf{S}^*$' and we must show that ϕ is a Π^1_2 formula. This is possible because of Lemma 8.3.5: we can restrict our attention to countable models $\mathfrak{M}$ and countable revision sequences S. Uncountable models or revision sequences can be mirrored by a countable elementary submodel (countable subsequence) making precisely the same formulas true.

Theorem 8.3.6 *Validity in* $\mathbf{S}^*$ *has precisely complexity* Π^1_2.

Proof: Validity in $\mathbf{S}^*$ is defined as follows: a sentence ϕ is valid in $\mathbf{S}^*$ if and only if for all classical models $\mathfrak{M}$ and all recurring hypotheses h, ϕ is true in $\mathfrak{M} + h$. In order to model this definition, we introduce new predicates: the two-place predicate $Ref(h, S)$ denotes the relation 'h is a reflexive hypothesis in the revision sequence S'. Furthermore, let $Rev(S, \mathfrak{M}, D)$ be the relation 'S is a revision sequence relative to a model $\mathfrak{M}$ and the (circular) definition D'. Now we can formalize validity (relative to a given circular definition D) in the system $\mathbf{S}^*$ as follows:

$$(1)\quad \forall\mathfrak{M}\forall h\forall S(Rev(S, \mathfrak{M}, D)\ \wedge\ Ref(h, S)\ \rightarrow\ \mathfrak{M} + h \models \phi)$$

A problem arises because of the unrestricted quantification over arbitrary models. If $\mathfrak{M}$ is uncountable, then the validity relation in our formalization (1) cannot be of any complexity class of the projective (analytical) hierarchy. Precisely at this point we can apply Lemma 8.3.5: it is possible to find an elementary submodel $\mathfrak{M}'$ of $\mathfrak{M}$ which is countable. Because of this fact, we can reconsider the definition of validity in $\mathbf{S}^*$ and reformulate it as follows:

$$(2)\quad \forall\mathfrak{M}\forall h\forall S\forall\alpha : (W(\alpha)\ \wedge\ Rev(S, \mathfrak{M}, D, \alpha)\ \wedge\ Ref(h, S, \alpha)\ \rightarrow\ \mathfrak{M} + h \models \phi)$$

First, we need to clarify the usage of the predicate W. The expression $W(\alpha)$ denotes the relation 'α is a well-ordering'. It is obvious that the property to be a well-order is at most a Π^1_1 relation. We need this relation W because we want to speak on the one hand about the length of countable revision sequences, and on the other hand about the existence of an ordinal $\beta < \alpha$, such that h is β-reflexive. Now, we have to determine the complexity of the relation 'α is a countable ordinal'. By definition the relation 'α is a countable ordinal' corresponds to the possible well-orderings of the natural numbers. With a result in descriptive set theory (using the fact that the relation 'α is a well-order' is Π^1_1), we know that that the relation 'α is a countable ordinal' is also Π^1_1.[12]

[12] Cf. [Mo80], 4A.2, p.192/193.

Furthermore, we are able to express $Rev(S, \mathfrak{M}, D, \alpha)$ arithmetically in some second-order parameters because of the following defining formula:

$$\begin{aligned}(3) \qquad Rev(S, \mathfrak{M}, D, \alpha) \;\Leftrightarrow\; & [\forall\gamma(\gamma = \beta + 1 \;\rightarrow\; S_\gamma = \delta(S_\beta)) \\ & \wedge\; \forall\lambda((\neg\exists\beta : \beta + 1 = \lambda) \\ & \rightarrow\; S_\lambda \text{ coheres with } S \restriction \lambda)]\end{aligned}$$

Notice that the operation induced by δ and the coherence condition are arithmetical relations. Finally, the relation 'h is a reflexive hypothesis in a revision sequence S' can be expressed as follows:

$$Ref(h, S, \alpha) \;\Leftrightarrow\; \exists\beta(\beta < \alpha \;\wedge\; h = S_0 = S_\beta)$$

Obviously, the above relation is arithmetical in the second-order parameters h and S.

The most complex relation in (2) is the validity relation relative to the ground model $\mathfrak{M}$ and the recurring hypothesis h. To show that validity (satisfaction) in 'nice' countable structures (we mean essentially acceptable structures in the sense of [Mo74] in this context) is of complexity Δ^1_1 is a tedious and sophisticated work. The idea is that the 'nice' countable structure $\mathfrak{M}$ has an elementary coding scheme which allows us to represent validity as the minimal fixed point of a positive formula. Using sequences, we code whether a formula is a negation of a formula, an atomic formula, or an existentially quantified formula etc. and show that this can be represented as a positive inductive definition. Then, this can be used to show that validity is an inductive relation (and therefore is a Π^1_1 relation). The converse, namely that validity is also a Σ^1_1 relation, follows easily by a trivial manipulation of the validity relation. Together we get that validity in our 'nice structures' is a Δ^1_1 relation. Now, we see that the definition of validity can be modeled using at most Π^1_1 relations and we can reduce everything to the countable case by the Löwenheim-Skolem theorem. Π^1_1 relations can be (strongly) defined in revision theories, therefore we have a Π^1_2 formula that defines validity in $\mathbf{S}^*$. Together with Corollary 8.3.4 we can conclude the claim of the theorem. q.e.d.

Remark 8.3.2 (i) Theorem 8.3.6 shows that the expressive power of $\mathbf{S}^*$ is quite strong. $\mathbf{S}^*$ is a proper extension of inductive and coinductive definitions, including additionally Δ^1_2 and Π^1_2 predicates.

(ii) What can be said about the complexity of $\mathbf{S}^\#$? It turns out that the complexity of the semantical system $\mathbf{S}^\#$ is also Π^1_2. This is easy to see, because the whole argumentation of Theorem 8.3.6 can be also used for the $\mathbf{S}^\#$ case. The only difference between the two systems is a first-order quantification over individual variables, specifying the finitely many further revisions after the α-th revision for $\alpha \in ORD$. That quantification does not create any problems.

In the last subsection, we shall consider additional aspects of the complexity theory of revision theories. Whereas in this section, the complexity of validities was our primary concern, we examine in the following section which relations are definable in Gupta-Belnap systems.[13] This is the usual way to address questions of complexity, as for example in the case of inductive definitions. Therefore, in the following we will ask which sets of ω can be defined in the semantical system $\mathbf{S}^*$ in the sense that there is a unique recurring hypothesis of a certain complexity in every possible revision sequence (relative to given definition).

8.4 Definability in Gupta-Belnap Systems

Whereas in the above Section 8.3 we examined the complexity of validities in $\mathbf{S}^*$ (and implicitly in $\mathbf{S}^\#$), in this section we shall discuss the properties of the semantical systems $\mathbf{S}^*$ and $\mathbf{S}^\#$ concerning the complexity of definable subsets. Notice that from the fact that validity in $\mathbf{S}^*$ is a Π^1_2 relation, we cannot deduce the complexity of an arbitrary circularly defined relation. Or in other words: from the knowledge of the complexity of validities one cannot deduce the complexity of the definable subsets of the domain ω. Obviously, a natural question of generalized definition theory is: Which relations (i.e. subsets of a given domain) can be defined using Gupta-Belnap systems? This section is devoted to this question.

First of all, we will define the notion 'X is definable in $\mathbf{S}^*$' precisely.[14]

Definition 8.4.1 *A subset $X \subseteq \omega$ is revision theoretically definable if and only if there is a (first-order) circular definition D defining a predicate P, such that D has a unique reflexive hypothesis h in every revision sequence S and the extension of P is equal to the extension of X.*

Remark 8.4.1 (i) It seems to be the case that there is no alternative to the uniqueness requirement of the reflexive hypothesis in Definition 8.4.1. In order to give an argument for the choice of this definition, assume G is a predicate (or a set) defined by the following circular definition:

$$G(x) \;\Leftrightarrow\; A(x, G)$$

Assume further that there are two reflexive hypotheses h_1 and h_2 with the property $h_1 \neq h_2$. Then it holds for all x: $G(x) \;\Leftrightarrow\; h_1(x)$ and $G(x) \;\Leftrightarrow\; h_2(x)$. This cannot be a desired consequence because a set that is definable should be definable in a unique way. This is not guaranteed if we allow that more than

[13] In order to simplify matters, we will restrict our attention to subsets of ω.

[14] Our considerations will be quite general. As a consequence we will not distinguish between the definability of $\mathbf{S}^*$ and $\mathbf{S}^\#$. The important instance will be the concept of recurring hypotheses and this concept is independent of semantical systems. Also, we only care about the definable subsets of ω. This is in a certain sense a restriction and simplification, but not a dramatic one: to determine the complexity of subsets of ω is the most interesting case.

one unique reflexive hypothesis can determine the extension of G.

(ii) Assume again the premises of (i). One could think that a good definition for definable subsets of ω in $\mathbf{S}^*$ would be the following:

$X \subseteq \omega$ is revision theoretically definable if and only if

$$\exists D : X = \bigcup\{x \mid x \text{ is stable in some recurring hypothesis } h \text{ for } D\}$$

That would be a generalization of Definition 8.4.1. In a certain sense, this definition would specify precisely all elements of the domain that are stable in some recurring hypotheses. Notice first that Definition 8.4.1 as well as the above alternative definition do not contain crucial properties of $\mathbf{S}^*$ except the concept of reflexive hypotheses.
What can be said concerning the differences between Definition 8.4.1 and the alternative definition above? Intuitively, the alternative definition is more general in comparison with Definition 8.4.1. For example, consider Truth-teller-like constructions. Using the alternative definition it is easy to define a predicate that contains stably true sentences and additionally biconsistent sentences. Although the alternative definition is interesting we will not consider it in detail here.

(iii) What can be said about further alternative definitions? The following definition can count as a further conceivable possibility.

$X \subseteq \omega$ is revision theoretically definable if and only if

$$\exists D \exists h : X = h \text{ where } h \text{ is a recurring hypothesis for } D$$

This alternative is more general than the definition in (ii). For example, it is possible to define predicates that contain not only stably true and biconsistent sentences but also paradoxical sentences as well. Even though this concept of definability is interesting we will not examine its properties here.

(iv) Let us consider a further version of an alternative definition.

$X \subseteq \omega$ is revision theoretically definable if and only if

$$\exists D : X = \bigcap\{x \mid x \text{ is stable in some recurring hypothesis } h \text{ for } D\}$$

Here, only those collections of elements can count as definable subsets where every element is in every recurring hypothesis. That means that precisely the elements that are stable in every recurring hypothesis are in the extension of D. An easy calculation shows that it is possible to find a definition D', such that the unique recurring hypothesis of D' determines the intersection of all recurring hypotheses h of D. Therefore, nothing can be gained by such an alternative of Definition 8.4.1.

(v) In recent work by Philip Welch[15] and Benedikt Löwe,[16] a different definition of revision theoretic definability is proposed. The authors say that a set $X \subseteq \omega$ is $\mathbf{S}^*$-definable if and only if

$$n \in X \iff \mathbb{N} \models^* n \in \dot{x}$$

where $n \in \dot{x}$ means that n is stably true in some revision sequence S. From an intuitive point of view this means to specify the subsets X of ω, such that the expression $n \in X$ is valid in $\mathbf{S}^*$. Intuitively, the Π^1_2 sets should be the result, because the set of all validities is a Π^1_2 subset of ω. In [Wel∞c], precisely this is proven.

In this section, we work in given structures where an effective coding scheme is available. In the sense of [Mo74], we can work in acceptable structures (cf. [Mo74], Chapter 5). Roughly speaking, one can consider standard arithmetic as the underlying structure of the whole consideration.

Now, we show that every inductive definition is revision theoretically definable. This is not a surprising fact when we reconsider the work in the above section, in particular when when we recall that we have already proven that fact in Lemma 8.3.2 (even though not explicitly).

Lemma 8.4.2 *Every Π^1_1 set X is revision theoretically definable.*

Proof: Assume X is a subset of ω. Recall the proof of Lemma 8.3.2. In the proof of this theorem, we constructed two circular definitions for the predicates G and H, such that the unique reflexive hypothesis for H was the interpretation of the minimal fixed point of an arbitrarily chosen (but fixed) inductive definition. Because of the fact that every inductive definition is precisely a Π^1_1 definition (compare Theorem 7.3.7), there is a circular definition that has a unique reflexive hypothesis h which is equal to the extension of X. Because this is the literal definition of revision theoretical definability, the claim of the theorem follows immediately. q.e.d.

It is clear that the concept "a set X is revision theoretically definable" is closed under negation. Therefore, we have the following result claiming that every coinductive relation is also definable.

Corollary 8.4.3 *Every Σ^1_1 set is revision theoretically definable.*

Proof: This is a trivial consequence of Lemma 8.4.2 because inductive sets that are revision theoretically defined are closed under negation (as can be seen

[15] Cf. [Wel∞c]

[16] Cf. [LöWe∞]

using Corollary 8.3.3). q.e.d.

We have shown that every inductive and every coinductive definition can be revision theoretically defined. We use the results of the above section in order to show that every Π^1_2 set is revision theoretically definable. In particular, we will use the property that every Π^1_2 set can be recursively embedded into the set $\{\phi \mid \phi \text{ is valid on } D \text{ in } \mathbf{S}_i\}$ (i.e. we apply the same idea as in Lemma 8.3.1). We shall show that for every finite list of positive elementary definitions in $\mathbf{S}_i$, we can define the set of Gödel numbers of the valid formulas in $\mathbf{S}_i$ on D.

Before we can do that, we need a further Lemma that shows that the relation "the countable structures a and b are isomorphic" is revision theoretically definable.

Lemma 8.4.4 *Let $\mathcal{A}$ and $\mathcal{B}$ be two countable structures. Then the relation*

$$R(\mathcal{A},\mathcal{B}) \textit{ if and only if } \mathcal{A} \cong \mathcal{B}$$

is revision theoretically definable.

Proof: Assume $\mathcal{A}$ and $\mathcal{B}$ are two countable structures. Assume further that both structures are isomorphic. Then, in every infinite Ehrenfeucht-Fraïssé game, player II has a winning strategy and in every infinite game in which player II has a winning strategy, the two structures are isomorphic (compare Theorem 7.3.9). The relation to have a winning strategy for player II for all infinite games (w.r.t. the given structures $\mathcal{A}$ and $\mathcal{B}$) can be expressed as follows (where $\Game(\mathcal{A},\vec{d},\mathcal{B},\vec{e})$ denotes an infinite game between player I and player II played in $\mathcal{A}$ and $\mathcal{B}$):

$$\begin{array}{ll}\forall G_n(\mathcal{A},\vec{x},\mathcal{B},\vec{y}): & (\forall n \in \omega : (x_1x_2...x_n \equiv y_1y_2...y_n) \ \rightarrow \\ & [(\forall x^* \exists y^* (x_1x_2...x_nx^* \equiv y_1y_2...y_ny^*))\wedge \\ & (\forall y^* \exists x^* (x_1x_2...x_nx^* \equiv y_1y_2...y_ny^*))])\end{array}$$

The representation is essentially a Π^1_1 (inductive) formula with a second order quantification over infinite games $G_n(\mathcal{A},\vec{d},\mathcal{B},\vec{e})$. According to Lemma 8.3.2, there are circular definitions for predicates G and H, such that $\langle x,y\rangle \in R$ iff $\langle x,y\rangle$ is in the extension of a unique reflexive hypothesis h for H where H determines the relation defined by the above formula. Conclude: $R(\mathcal{A},\mathcal{B})$ iff $\mathcal{A} \cong \mathcal{B}$ is revision theoretically definable. q.e.d.

Remark 8.4.2 Notice that the direct logical representation of equivalence between two structures is not appropriate here. The following formula describes the property of two structures to be equivalent.

$$\begin{aligned} R(\mathcal{A},\mathcal{B}) \quad \Longleftrightarrow \quad & \exists f : A \longrightarrow B \ \wedge \ f \text{ bijective} \wedge \\ & \forall R : R(a_1, a_2, ..., a_n) \ \leftrightarrow \ R(f(a_1), f(a_2)), \ldots, f(a_n)) \end{aligned}$$

Notice that it is not immediately clear that the relation defined by this formula can be defined using revision theoretic techniques.

The following theorem proves that every Π^1_2 relation is revision theoretically definable. In the proof, we use quite similar techniques and ideas as in the proof of Theorem 8.3.6 where we showed that validity in $\mathbf{S}^*$ has complexity Π^1_2.

Lemma 8.4.5 *Every Π^1_2 subset of ω is revision theoretically definable.*

Proof: Assume A is an arbitrarily chosen Π^1_2 subset of ω. We know by Lemma 8.3.1 that it holds (with the appropriate specifications of Q and H):

$$\forall n \in \omega : (n \notin A) \iff \models_i PA^- \ \wedge \ SLO \ \wedge \ \forall x(Hx) \ \rightarrow \ \neg Q(n)$$

In the above expression, H and Q are defined as in Lemma 8.3.1, i.e. we are working in an extended language L^+ of arithmetic: $L^+ = \{0,', +, \cdot, <\} \cup \{Q, Y, H\}$. We need a way to define the following set Z revision theoretically:

$$Z \ = \ \{n \in \omega \mid \models_i PA^- \ \wedge \ SLO \ \wedge \ \forall x(Hx) \ \rightarrow \ \neg Q(n)\}$$

First, notice that the following statements are equivalent:

$$\begin{aligned} & \models_i \ PA^- \ \wedge \ SLO \wedge \ \forall x(Hx) \ \rightarrow \ \neg Q(n) \\ \iff \quad & \forall \mathfrak{M} : (\mathfrak{M} + H + Q + Y \models PA^- \ \wedge \ SLO \ \wedge \ \forall x(Hx) \ \rightarrow \ \neg Q(n)) \\ \iff \quad & \forall \mathfrak{M} : \mathfrak{M} \cong \mathbb{N} \ \rightarrow \ \mathfrak{M} + H + Q + Y \models \neg Q(n) \\ \iff \quad & \forall \mathfrak{M} : \mathfrak{M} \cong \mathbb{N} \ \rightarrow \ \mathfrak{M} + Y \models_i \neg Q(n) \end{aligned}$$

The last formula is essentially a Π^1_1 formula, but we have an unrestricted quantification over (possibly uncountable) models. We have to use a similar technique as in Lemma 8.3.5 to guarantee that it is possible to work in countable structures. In other words: we need an application of the Löwenheim-Skolem Theorem to ensure that we can work with a restricted quantifier, quantifying only over countable models. Because there is no additional complication for applying the Löwenheim-Skolem Theorem (like in Lemma 8.3.5), we omit a proof. Then, we get the countable version of the last equivalence of the above reasoning:

$$\forall \mathfrak{M}_{count} : (\mathfrak{M}_{count} \cong \mathbb{N} \ \longrightarrow \ \mathfrak{M}_{count} + Y \models_i \neg Q(n))$$

By using Lemma 8.4.4, we know that the property of two structures to be isomorphic is definable revision theoretically. We now consider the satisfaction relation. It is well known that the satisfaction relation in arithmetic (more generally in acceptable structures that include arithmetic) is a Δ^1_1 relation.[17] Because we work essentially in a model that is isomorphic to $\mathbb{N}$, the satisfaction relation is still a Δ^1_1 relation. This is also true, when we expand our model by inductive relations H and Q provided the extensions of H and Q are specified. The proof for that claim is similar to the standard proof that satisfaction in acceptable structures is Δ^1_1. We can conclude that the formula

$$\forall \mathfrak{M}_{count} : (\mathfrak{M}_{count} \cong \mathbb{N} \ \longrightarrow \ \mathfrak{M}_{count} + Y \models_i \neg Q(n))$$

is a Π^1_1 expression (therefore defining an inductive relation) with revision theoretically definable parameters $\models, \cong, \mathbb{N}$, and Q. Because inductive relations can be defined revision theoretically (compare Lemma 8.4.2), we have the following fact:

$$Z = \{n \in \omega \mid \models_i PA^- \ \wedge \ SLO \ \wedge \ \forall x(Hx) \ \rightarrow \ \neg Q(n)\}$$

is revision theoretically definable. Therefore, the set $A \subseteq \omega$ is revision theoretically definable by taking the complement. Because A was arbitrarily chosen, this suffices for the proof of the theorem. q.e.d.

Remark 8.4.3 (i) We were not fully explicit in Theorem 8.4.5. We did not show that satisfaction in acceptable structures is a Δ^1_1 relation. To do that we need to specify a particular coding scheme for formulas in our language and we need to mirror the proof in [Mo74], Chapter 5 for our purposes. Hence, we need to show that adding finitely many (in fact only two) inductive definitions to our system does not change the complexity of satisfaction. But this is quite easy to prove, because the extensions of the additional inductive relations are fixed.

(ii) Theorem 8.4.5 does not provide us with an upper bound of the complexity of definable relations (sets) in Gupta-Belnap systems. That holds, because the notion "revision theoretically definable" is closed under negation (unlike the complexity class Π^1_2 itself that is clearly not closed under taking the complement). The next Lemma 8.4.6 tells us that even Σ^1_2 relations are revision theoretically definable.

Lemma 8.4.6 *Every Σ^1_2 subset A of ω is revision theoretically definable.*

Proof: The proof is similar to Corollary 8.3.3 and uses the closure property of negation of revision theoretically defined sets: Assume A is a Σ^1_2 subset of ω. Construct the Π^1_2 set A^C (complement of A). By using the same technique as in Theorem 8.4.5, we can show that A^C can be defined by a revision theoretic definition D'. In other words: the unique reflexive hypothesis h' of D' gives us the extension of A^C. More precisely, define a predicate H' by the unique

[17] Compare [Mo74], Chapter 5 for further information concerning the properties of acceptable structures.

reflexive hypothesis h' and calculate the extension of A by the predicate H according to the following formula:

$$H(x) \Leftrightarrow \neg H'(x)$$

Then it holds:

$$x \in A \Leftrightarrow x \notin A^C \Leftrightarrow \neg H'(x) \Leftrightarrow H(x)$$

Because H' has a unique reflexive hypothesis h', H has also a unique reflexive hypothesis h. Conclude: h defines A. q.e.d.

The next Lemma specifies an upper bound for the complexity of the concept of revision theoretic definability.

Lemma 8.4.7 *Revision theoretic definability is at most of complexity Δ^1_3.*

Proof: The formal representation of Definition 8.4.1 can be specified as follows.

$$\begin{aligned} &X \subseteq \omega \text{ is revision theoretic definable if and only if} \\ &\qquad \exists D \forall S \forall \mathfrak{M}_{count} \exists! h : [(D \text{ circular definition}) \\ &\qquad \wedge\ (Rev(S, \mathfrak{M}_{count}, D, \alpha) \rightarrow Ref(h, S, \alpha)) \\ &\qquad \wedge\ (\forall x : x \in X \leftrightarrow x \in h)] \end{aligned}$$

The above formula is a Σ^1_3 formula with several parameters in it. Notice that the property to be a circular definition is elementary. It is clear that the same techniques we used in Section 8.3 Theorem 8.3.6 can be applied in the above considerations in order to make sure that we do not need an unrestricted quantification over arbitrary models. Although we are not completely explicit in the above formula (for example, we do not examine the properties of the quantifier $\exists!$ and we do not mention that D can contain more than one circular definition) it is obvious that the above Σ^1_3 formula is precisely the definition of the property to be revision theoretical definable.

On the other hand, it is also possible to determine revision theoretic definability as follows.

$$\begin{aligned} &X \subseteq \omega \text{ is revision theoretic definable if and only if} \\ &\qquad \forall \mathfrak{M}_{count} \exists D \forall S \forall h \forall h' : [(D \text{ circular definition}) \\ &\qquad \wedge\ (Rev(S, \mathfrak{M}_{count}, D, \alpha) \wedge Ref(h, S, \alpha))] \\ &\qquad \wedge\ (Rev(S, \mathfrak{M}_{count}, D, \alpha) \wedge Ref(h', S, \alpha) \rightarrow h = h') \\ &\qquad \rightarrow (x \in X \leftrightarrow x \in h) \end{aligned}$$

Easy logical considerations show that the above formula precisely defines the property to be revision theoretic definable. This fact together with the above representation implies that revision theoretic definability is at most of complexity Δ^1_3. q.e.d.

Remark 8.4.4 (i) We simplified the presentation because the techniques are quite similar to the techniques used in Section 8.3.

(ii) Lemma 8.4.7 gives us an upper bound for the complexity of revision theoretic definability. On the other hand, the Lemmas 8.4.5 and 8.4.6 determine a lower bound of the complexity of revision theoretic definability. The following Conjecture 8.4.8 tries to make clear the intuitions of a general result concerning that problem.

Conjecture 8.4.8 *The property of being revision theoretic definable is precisely of complexity Δ^1_3.*

Remark 8.4.5 The conjecture that revision theoretic definability is precisely a Δ^1_3 statement seems at least the intuitively plausible candidate. If there is a possibility to categorize the complexity of revision theoretic definability into the projective hierarchy at all, then it would be rather implausible if the complexity was $\Sigma^1_3 \cup \Pi^1_3$. A natural result would be that the complexity is Δ^1_3.

We finish this chapter with these remarks. In the next chapter, we will examine possible applications of revision theories in different fields ranging from a theory of truth to considerations to define a non-well-founded set theoretic universe.

8.5 History

Revision theories as discussed in this chapter originated with four papers in the year 1982, namely [Gu82, Be82, He82a, He82b]. The most important difference between the systems discussed in these papers was the question of how an intuitively correct limit rule for the revision process can be established. The motivation for developing an alternative theory for circularity was the Liar paradox and pathological sentences in natural language in general and the inappropriate treatment of Tarski's and Kripke's account for some of the empirical data. Later, other applications were found by examined by Aldo Antonelli (compare [An94b, An94c]) and by André Chapuis (work in progress). With the monograph [GuBe93], Gupta and Belnap provided a book including most of the important results of revision theories. In this work, the question which complexities the semantical systems $\mathbf{S}^*$ and $\mathbf{S}^\#$ do have was raised for the first time (as far as the author knows). A further question was whether these systems are recursively axiomatizable or not. The result of the non-axiomatizability of these systems and the fact that validity of these systems is at least Π^1_2 are due to [Kr93]. Aldo Antonelli proved in [An94a] and [An98] the result that this upper bound is optimal. A further reference for results (and the important McGee Theorem) was formulated in [Mc91]. The questions concerning the definable subsets of ω was raised by the author. In [Wel∞c] and [LöWe∞], some answers to a particular concept of revision theoretic definability were proven. The facts concerning the definability of subsets of ω as defined in Definition 8.4.1 are new here.

Chapter 9

Applications of Gupta-Belnap Systems

In this chapter, we will examine how revision theories can be used in order to model different circular phenomena. Originally, Gupta-Belnap systems were developed to give an appropriate approach for a theory of truth. The motivation arose from some unsatisfactory and counterintuitive results and consequences of Kripke's fixed point approach.[1] Because of these origins of revision theories, most work that was done in the field of applying revision theories to circular phenomena deals with the attempt to develop an intuitively correct theory of truth. Although it is relatively clear that revision theories can also be used in other areas that involve circularity, it is a relatively new development in the history of this framework to try to model other phenomena as well.[2]

We saw in Part II that Kripke's basic idea was to change the underlying logic and to assign non-classical truth values to pathological sentences. Although this account gives a lot of appropriate results for a theory of truth it does not tell us anything about the intrinsic behavior of circularity. In a certain sense, circularity is still banned: We simply block paradoxes by assigning non-classical truth values to these sentences. Gupta's idea (as well as Herzberger's and Belnap's idea) is to affirm circularity as a respectable phenomenon that is included in different forms of reasoning and conceptualizations. Revision theories evolve circularity in a sequence of hypotheses (based on a given revision rule), in order to observe the behavior of this sequence and evaluate the whole sequence of hypotheses in an appropriate semantical system. All this is done in a classical logic. What is needed is an enlarged version of a semantical system that does not yield inconsistencies if one defines a circular predicate. The differences between pathological sentences and false sentences is not captured on a logical level, but on the level of the behavior of extensions of the predicate in the revision process. Keeping this in mind, it is clear that revision theories differ quite significantly from accounts using

[1] Compare Chapter 6 for further information.

[2] Currently André Chapuis is working on applications of Gupta-Belnap systems to model belief nets and their revision theoretic properties (personal communication with Anil Gupta). To use revision theories to develop a theory of knowledge representation is (as far as the author knows) not developed yet.

monotone logics and a fixed point construction.

We will consider certain properties of possible applications of revision theories to a theory of truth, to set theory, and to mathematics in general in this chapter. We do not claim to be complete or to be completely explicit in our development. The aim is to present an overview and to stimulate other researchers to do further work.

9.1 The Modeling of Pathological Sentences

In this section, we assume that our domain D is given by the two classical truth values T and F. Provided a language L is given, a revision rule ρ is a mapping induced by a (circular) definition and specified as follows:

$$\rho : D^{Sent_{L^+}} \longrightarrow D^{Sent_{L^+}}$$

We will consider sentences of natural language and model and evaluate them with revision theoretic means. As is generally accepted by authors working in Gupta-Belnap systems (and in fact as a general basis of the whole account) the revision rules ρ are crucially based on Tarski's biconditionals: One of the motivations for revision theories was the attempt to find a theory based on classical logic that preserves Tarski's biconditionals.[3]

First, we need to introduce the evaluation scheme for pathological sentences more precisely. The semantical evaluation of revision sequences in this section is simply a special form of the evaluation of ordinary revision sequences as developed in Chapter 8.

Definition 9.1.1 *Let $L^+ = L \cup \{\mathbf{T}\}$ be a given language extended by a truth predicate $\mathbf{T}$. Let $\mathfrak{M}$ be a classical ground model. The set $\{T, F\}^{Sent_{L^+}}$ is defined as follows:*

$$\bigcup \{f : \{\phi : \phi(x_1, x_2, ..., x_n) \in Form_{L^+}\} \times D^n \longrightarrow \{T, F\}\}$$

We say that relative to $\mathfrak{M}$ the expression $\langle \phi, \vec{d} \rangle(h) = T$ if and only if $\phi(\vec{d})$ is true in $\mathfrak{M} + h$.

Remark 9.1.1 An equivalent formulation of the phrase $\phi(\vec{d})$ is true in $\mathfrak{M} + h$ (and perhaps for many readers a more intuitive one) can be stated as follows:

$$\phi(\vec{d}) \text{ is true in } \mathfrak{M} + h \;\Leftrightarrow\; [[\phi(\vec{d})]]^{\mathfrak{M}+h} = T$$

We will use the latter formulation as an abbreviation of the formulation given in Definition 9.1.1, because it is more common and easier to understand.[4]

[3] As a matter of fact, revision theories were developed by the idea that a theory of truth need not only preserve Tarski's biconditionals, but must also be governed by these biconditionals.

[4] Definition 9.1.1 goes back to [GuBe93].

The evaluation of the sentences in the revision sequence is based on a classical two-valued logic. This is an important difference to the fixed point approaches we examined in Part II of this work. An evaluation sequence $\langle S_\phi \rangle$ of a sentence $\phi \in Sent_{L^+}$ relative to a given revision rule ρ and a revision sequence S of hypotheses is defined as follows:

$$\langle S_\phi \rangle \;=\; \langle [[\phi(\vec{d})]]^{\mathfrak{M}+\rho^\beta(h)} \rangle_{\beta<length(S)}$$

The introduction of an evaluation sequence $\langle S_\phi \rangle$ of a single sentence ϕ makes it easier to study the behavior of a single pathological sentence. It is not a crucial concept of the theory, but merely a possibility to refer to single sentences instead of a whole collection of sentences.

We can define validity in $\mathbf{S}^*$ as a special case of the general Definition 8.1.4(i).

Definition 9.1.2 *We say that a sentence ϕ is valid in $\mathfrak{M}$ if and only if the following condition holds:*

$$\mathfrak{M} \models^* \phi \iff \forall h : Ref_D(h) \rightarrow [[\phi(\vec{d})]]^{\mathfrak{M}+h} = T$$

It is easy to see that Definition 9.1.2 is precisely the claim in Definition 8.1.4(i) specified appropriately in order to make it easier to apply it to natural language phenomena.

In Kripke's account, it is possible to distinguish two forms of pathological sentences: paradoxical sentences (for example the Liar sentence or a Liar circle) are sentences that are ungrounded and unstable relative to all possible hypotheses h. Biconsistent sentences (an example is the Truth-teller sentence) are sentences that are ungrounded but are stably true or stably false dependent on the initial hypothesis. Although these distinctions can clarify matters and show the different behavior of sentences, Gupta-Belnap systems are able to make finer distinctions.[5] The idea of these further distinctions is to distinguish sentences dependent on their behavior in different evaluation sequences. If a sentence ϕ is unstable in every possible evaluation sequence, then ϕ shows a different behavior in comparison to a sentence ψ that is unstable in certain evaluation sequences but stable in others.[6] The following definition introduces these further distinctions between sentences.

Definition 9.1.3 *Assume a language $L^+ = L \cup \{\mathbf{T}\}$ and a classical ground model $\mathfrak{M}$ are given.*
(i) A sentence $\phi(\vec{d})$ is called uniformly true, if for all recurring hypotheses h in an arbitrary evaluation sequence $\langle S_\phi \rangle$ it holds: $[[\phi(\vec{d})]]^{\mathfrak{M}+h} = T$.
(ii) A sentence $\phi(\vec{d})$ is called uniformly false, if for all recurring hypotheses h

[5] The following distinctions were first introduced by Yaqūb in his book [Ya93].

[6] A sentence ϕ is valid in the semantical system $\mathbf{S}^*$, if ϕ is evaluated stably true in all possible evaluation sequences.

in an arbitrary evaluation sequence $\langle S_\phi \rangle$ *it holds:* $[[\phi(\vec{d})]]^{\mathfrak{M}+h} = F$.
(iii) A sentence $\phi(\vec{d})$ *is called unstable, if for all recurring hypotheses h in an arbitrary evaluation sequence* $\langle S_\phi \rangle$ *it holds:* $\phi(\vec{d})$ *is unstable in* $\langle S_\phi \rangle$.
(iv) A sentence $\phi(\vec{d})$ *is called* TF*-capricious, if for all recurring hypotheses h in an arbitrary evaluation sequence* $\langle S_\phi \rangle$ *it holds: either* $[[\phi(\vec{d})]]^{\mathfrak{M}+h} = T$ *or* $[[\phi(\vec{d})]]^{\mathfrak{M}+h} = F$.
(v) A sentence $\phi(\vec{d})$ *is called* TN*-capricious, if for all recurring hypotheses h in an arbitrary evaluation sequence* $\langle S_\phi \rangle$ *it holds: either* $[[\phi(\vec{d})]]^{\mathfrak{M}+h} = T$ *or* $\phi(\vec{d})$ *is unstable in* $\langle S_\phi \rangle$.
(vi) A sentence $\phi(\vec{d})$ *is called* FN*-capricious, if for all recurring hypotheses h in an arbitrary evaluation sequence* $\langle S_\phi \rangle$ *it holds: either* $[[\phi(\vec{d})]]^{\mathfrak{M}+h} = F$ *or* $\phi(\vec{d})$ *is unstable in* $\langle S_\phi \rangle$.
(vii) A sentence $\phi(\vec{d})$ *is called* TFN*-capricious, if for all recurring hypotheses h in an arbitrary evaluation sequence* $\langle S_\phi \rangle$ *it holds: either* $[[\phi(\vec{d})]]^{\mathfrak{M}+h} = T$*, or* $[[\phi(\vec{d})]]^{\mathfrak{M}+h} = F$*, or* $\phi(\vec{d})$ *is unstable in* $\langle S_\phi \rangle$.

Remark 9.1.2 (i) The above definition captures the idea that pathological sentences can be categorized in a variety of different classes, dependent on their behavior. In fixed point approaches, there is the possibility to get something quite similar, if one allows to consider not only single fixed points but sets (collections) of fixed points as well. For example, the Truth-teller sentence has the truth value N in the minimal fixed point, in another fixed point the truth value T, in another one the truth value F. Therefore, even in fixed point theories one can express some of the presented classifications in Definition 9.1.3.

(ii) Definitions 9.1.3(i) and (ii) correspond to the grounded sentences in fixed point approaches. These cases are not very interesting, because these sentences are already captured by the ground model plus Tarski's biconditionals. Condition (iii) corresponds to the paradoxical sentences whereas condition (iv) corresponds to biconsistent sentences. Conditions (v)-(vii) are not explicitly mentioned in fixed point accounts, but, as was mentioned above, they can be modeled in fixed point theories by considering appropriate subsets of fixed points.

(iii) The usage of the symbol N for an unstable behavior of a sentence in an evaluation sequence is slightly misleading, because we work in a classical logic without a third truth value. N does not represent this third truth value (as it was the case in the fixed point approach). N denotes the fact that in an evaluation sequence a certain sentence is neither stably true nor stably false. From this perspective the usage of the symbol N can be justified although revision theories are based on a classical bivalent logic.

From the basic definitions of revision theories it is clear that ordinary sentences that do not contain the truth predicate do not cause any problems in the framework. Their extensions are given by the ground model and their semantics is determined by classical model theoretic tools. Hence, these examples are relatively uninteresting and we will not consider those in detail. We will examine some easy (but more interesting) examples of sentences that show a non-classical behavior in revision theories.

Example 9.1.3 (i) Assume a language $L^+ = L \cup \{\mathbf{T}\}$ and a classical ground model $\mathfrak{M}$ are given. Consider the following two sentences:[7]

(1) $\phi \rightarrow \mathbf{T}(\phi)$
(2) $\phi \wedge \neg\mathbf{T}(\phi)$

(1) is one direction of an instance of Tarski's biconditional. Therefore, (1) is uniformly true. It is easy to see that the recurring hypotheses of (1) are the ones that assign the truth value T to (1). A similar consideration assures that (2) is uniformly false, because the recurring hypotheses h for (2) are the ones that satisfy $[[(\phi \wedge \neg\mathbf{T}(\phi))]]^{\mathfrak{M}+h} = F$.

What happens if we substitute for ϕ the Liar sentence λ in (1)? Differently to the treatment in fixed point theories where the sentence $\lambda \rightarrow \mathbf{T}(\lambda)$ is evaluated as neither true nor false (using Kleene's three-valued logic), the newly created sentence remains uniformly true in revision theories. That is a direct consequence of the fact that Tarski's biconditionals are assumed to govern the concept of truth. Later (compare Subsection 10.1.4) we will consider a certain type of criticism that was uttered because of the fact that revision theories preserve tautologies.

(ii) We consider the ordinary Liar sentence (3):

(3) $\lambda \ = \ \neg\mathbf{T}(\lambda)$

If we start the revision theoretic analysis with the hypothesis $h(\lambda) = T$, then the next revision step results in the evaluation $\rho(h(\lambda)) = F$. The next revision yields $\rho(\rho(h(\lambda))) = T$. This behavior continues through all ordinals. A similar instability arises if we start with the hypotheses $h(\lambda) = F$. This shows that both hypotheses are recurring. In total, we get the result that for all recurring hypotheses in every evaluation sequence $\langle S_\lambda \rangle$ the sentence $\lambda \ = \ \neg\mathbf{T}(\lambda)$ is unstable in $\langle S_\lambda \rangle$.[8]

(iii) Consider the ordinary Truth-teller sentence (4).

(4) $\phi \ = \ \mathbf{T}(\phi)$

[7] In the following, we will not represent the code of a sentences ϕ if this sentence is in the range of the truth predicate $\mathbf{T}$. Hence, we write $\mathbf{T}(\phi)$ instead of $\mathbf{T}(\ulcorner\phi\urcorner)$.

[8] We haven't considered the behavior in limit stages, yet. This is simple to check: whatever our new initial hypothesis is in a limit stage, the revision process never becomes stable.

A similar examination as in (ii) shows that the recurring hypotheses remain stably T or stably F in every evaluation sequence $\langle S_\phi \rangle$ dependent on the choice of the initial hypothesis. That means that the sentence $\phi = \mathbf{T}(\phi)$ is TF-capricious as it is expected. The evaluation of this sentence is completely independent of the choice of the limit rules, because the evaluation becomes stable after one revision.

(iv) It should be mentioned that all sentences not containing the truth predicate $\mathbf{T}$ are evaluated according to the ground model $\mathfrak{M}$. In an arbitrary revision sequence S with an initial hypothesis h, they become stable after at most one revision.

We continue the overview of examples by examining more complex ones. First, we will consider the Liar circle of length 2. A Liar circle should show a similar behavior like the ordinary Liar sentence. This means that the set of sentences should be interpreted as unstable, i.e. all sentences should be unstable in all possible evaluation sequences.

Example 9.1.4 We represent the Liar circle (of length 2) as follows (we choose length 2 because it is the simplest non-trivial Liar circle).

$$\begin{aligned}(5)\ \phi &= \mathbf{T}(\psi)\\ \psi &= \neg\mathbf{T}(\phi)\end{aligned}$$

We represent the behavior of the Liar circle in the following tables. Because (5) consists of two sentences there are $2^2 = 4$ many possibilities for initial hypotheses (both sentences can be true or false and all combinatorial combinations are allowed). The first table shows the behavior of the evaluation sequence with the initial hypothesis $h(\phi) = T$ and $h(\psi) = T$.

	0	1	2	3	4	5	6	7	$\cdots$
$\phi = \mathbf{T}(\psi)$	T	T	F	F	T	F	F	T	$\cdots$
$\psi = \neg\mathbf{T}(\phi)$	T	F	F	T	F	F	T	F	$\cdots$

We start with the hypothesis that $h(\phi) = T$ and that $h(\psi) = T$. The following rows show how the hypothesis is revised by the revision process. The revision remains unstable through all ordinals. Let us consider the next table where we start with the following hypothesis: $h(\phi) = T$ and $h(\psi) = F$.

	0	1	2	3	4	5	6	7	$\cdots$
$\phi = \mathbf{T}(\psi)$	T	F	F	T	F	F	T	F	$\cdots$
$\psi = \neg\mathbf{T}(\phi)$	F	F	T	F	F	T	F	F	$\cdots$

Again the revision process is not stable (as it is expected). The next table shows the behavior of the revision process when we start with the hypothesis $h(\phi) = F$ and $h(\psi) = T$. Again we get an unstable sequence of hypotheses.

	0	1	2	3	4	5	6	7	$\cdots$
$\phi = \mathbf{T}(\psi)$	F	T	F	F	T	F	F	T	$\cdots$
$\psi = \neg\mathbf{T}(\phi)$	T	F	F	T	F	F	T	F	$\cdots$

The last table shows the behavior of the revision process if we start with the hypothesis $h(\phi) = F$ and $h(\psi) = F$.

	0	1	2	3	4	5	6	7	$\cdots$
$\phi = \mathbf{T}(\psi)$	F	F	T	F	F	T	F	F	$\cdots$
$\psi = \neg\mathbf{T}(\phi)$	F	T	F	F	T	F	F	T	$\cdots$

In total, we see that the Liar circle is unstable, because every hypothesis we choose for a limit stage yields an unstable revision process. This result corresponds to our intuitions. Additionally, one can see that the behavior of the revision is completely regular. We can conclude: The analysis of the Liar circle (of length 2) in revision theories is in accordance to our intuition.

There is another example we should consider for a moment: the Truth-teller circle. This example is interesting because of two features. First, the Truth-teller circle is an example for a TFN-capricious collection of sentences when analyzed using revision theoretic tools. Second, this analysis is intuitively not the correct one, because we would expect that the Truth-teller circle is TF-capricious. The sentences should be stable dependent on their initial hypothesis. We will consider this example more closely now.

Example 9.1.5 We represent the Truth-teller circle (of length 2) by the following two formulas:

$$(6)\quad \begin{aligned} \phi &= \mathbf{T}(\psi) \\ \psi &= \mathbf{T}(\phi) \end{aligned}$$

Similarly to Example 9.1.4, we have to check the possible evaluation sequences of the two sentences. We start with the initial hypothesis h, such that it holds: $h(\phi) = T$ and $h(\psi) = T$. The following table shows the behavior of the resulting evaluation sequence.

	0	1	2	3	4	5	6	7	$\cdots$
$\phi = \mathbf{T}(\psi)$	T	T	T	T	T	T	T	T	$\cdots$
$\psi = \mathbf{T}(\phi)$	T	T	T	T	T	T	T	T	$\cdots$

This is an intuitively correct behavior of the Truth-teller circle: if we start the revision process with the initial hypothesis h, such that $h(\phi) = T$ and $h(\psi) = T$, then ϕ and ψ remain stably true through the complete revision process. Consider the next table now. Our initial hypothesis is $h(\phi) = T$ and $h(\psi) = F$. We would like to get one or the other form of a stable behavior of the evaluation sequence, but unfortunately the sequence is essentially unstable as the following table shows.

	0	1	2	3	4	5	6	7	$\cdots$
$\phi = \mathbf{T}(\psi)$	T	F	T	F	T	F	T	F	$\cdots$
$\psi = \mathbf{T}(\phi)$	F	T	F	T	F	T	F	T	$\cdots$

A similar behavior can be observed, if we choose our initial hypothesis to be $h(\phi) = F$ and $h(\psi) = T$.

	0	1	2	3	4	5	6	7	$\cdots$
$\phi = \mathbf{T}(\psi)$	F	T	F	T	F	T	F	T	$\cdots$
$\psi = \mathbf{T}(\phi)$	T	F	T	F	T	F	T	F	$\cdots$

Finally the situation remains stable if our initial hypothesis is $h(\phi) = F$ and $h(\psi) = F$.

	0	1	2	3	4	5	6	7	$\cdots$
$\phi = \mathbf{T}(\psi)$	F	F	F	F	F	F	F	F	$\cdots$
$\psi = \mathbf{T}(\phi)$	F	F	F	F	F	F	F	F	$\cdots$

In total, we see that the Truth-teller circle is a TFN-capricious example, although this analysis is intuitively not correct.

Because of the problematic analysis of the Truth-teller circle when analyzed revision theoretically we add some remarks concerning the behavior of this example.

Remark 9.1.6 When considering Example 9.1.5 the question arises, what are the reasons for the inappropriate analysis of revision theories?[9] Notice that our tables show only the behavior of the evaluation sequences for finitely many revisions. The theory developed in Chapter 8 (especially the development of the semantical systems $\mathbf{S}^*$ and $\mathbf{S}^{\#}$) are based on revision sequences that have length ORD. If we do not work with finite revision sequences but with infinite ones, the natural question arises which hypotheses are appropriate for unstable elements at a limit stage. We adopted in Definition 8.1.1(ii) the limit rule that determines only the denotation of the stable sentences but states no requirements for the unstable elements.[10] Applied to our Example 9.1.5 with initial hypotheses $h(\phi) = T$ and $h(\psi) = F$ we can choose again the same initial hypothesis for the two unstable sentences at every limit stage. This choice is allowed by our limit rule and guarantees furthermore that the revision process shows the same behavior through all ordinals as in the finite case of the table above, provided we do not assign the same truth value to ϕ and ψ at a limit stage. The quite liberal choice of the limit rule is essentially the reason why we get the intuitively incorrect result of Example 9.1.5. Clearly, alternative choices like the Herzberger rule or the Belnap rule would not fix the problem either, because in the Herzberger case we are free to choose an interpretation of the unstable elements at limit stages from the very beginning. That would also result in an evaluation sequence where all sentences are unstable. In the Belnap case, we could choose as initial hypothesis h with $h(\phi) = T$ and $h(\psi) = F$ and the result would be again an unstable evaluation sequence. Both limit rules result in a similar behavior of the evaluation sequence. The situation changes when we require that for unstable elements all possible combinations of truth values need to be assigned in limit stages. Precisely this solution is proposed in [Ch96]. We postpone the discussion of his solution to Section 10.1.

A further example we will mention is the revision theoretic analysis of the Gupta puzzle. As was pointed out in Chapter 2 Example (7) this is an example where a discourse is involved and where fixed point theories cannot assign definite truth values. Whereas fixed point theories are facing a problem with this kind of example, revision theories provide a better result that is in accordance to ordinary common sense reasoning.

Example 9.1.7 In order to make things easier, we analyze a simplified version of the Gupta puzzle. The reference to the 'real world' in Chapter 2 Example (7) is not crucial for the correct reasoning. Similarly, the problems the Kripke account has with this example is based on the fact that there is no possibility to assign a truth value to the sentences referring to statements of other persons. We choose the following form of the Gupta puzzle emphasizing the crucial part of the discourse.

[9] In [Ch96] and [Ya93], one can find a further discussion of this problem of classical revision theory. Both authors propose a solution for this problem. Another source for further information concerning this point is [Ma96]. We discuss possible solutions in Section 10.1.

[10] Compare also Remark 8.1.1(ii) for further information concerning this point.

(7) A claims (ϕ_1) and (ϕ_2):
(ϕ_1) All claims of B are true.
(ϕ_2) At least one of the claims of B is false.
B claims (ψ):
(ψ) At most one of the claims of A is true.

We can formalize the discourse, such that the revision process in evaluation sequences can be calculated more easily. (8) is a representation of (7):

(8) $\phi_1 = \mathbf{T}(\psi)$
$\phi_2 = \neg\mathbf{T}(\psi)$
$\psi = \neg(\mathbf{T}(\phi_1) \wedge \mathbf{T}(\phi_2))$

The following table shows the behavior of the revision process (for the first finitely many revision steps) where ϕ_1, ϕ_2, and ψ are assumed to be true at the initial stage:

	0	1	2	3	4	5	6	7	...
$\phi_1 = \mathbf{T}(\psi)$	T	F	F	T	T	T	T	T	...
$\phi_2 = \neg\mathbf{T}(\psi)$	T	T	T	F	F	F	F	F	...
$\psi = \neg(\mathbf{T}(\phi_1) \wedge \mathbf{T}(\phi_2))$	T	F	F	T	T	T	T	T	...
$\mathbf{T}(\phi_1)$	T	T	F	F	T	T	T	T	...
$\mathbf{T}(\phi_2)$	T	T	T	T	F	F	F	F	...
$\mathbf{T}(\psi)$	T	T	F	F	T	T	T	T	...

We can see that the revision process stabilizes for the initial hypothesis $h(\phi_1) = T$, $h(\phi_2) = T$, and $h(\psi) = T$. (The respective denotations of the sentences $\mathbf{T}(\phi_1)$, $\mathbf{T}(\phi_2)$, and $\mathbf{T}(\psi)$ are calculated using Tarski's biconditionals.) It is easy to check that the evaluation remains stable for every possible initial hypothesis. We will not explicitly calculate the evaluation sequences for the 2^3 possible initial hypotheses.

Although our version of the Gupta puzzle is simplified, it is clear that the extension to the version of the puzzle presented in Chapter 2 Example (7) is straightforward and does not change anything in the revision process, because the truth values of the sentences referring to the 'real world' are captured by the given ground model. They are not dependent on a specific property of a particular Gupta-Belnap system.

Unfortunately, a slight modification of the classical Gupta puzzle causes problems for revision theories.[11] We will state the modified Gupta puzzle as a further example.

[11] Cf. [GuBe93] and [Ch96].

Example 9.1.8 The modified Gupta puzzle can be formalized as follows:[12]

(9) $\phi_1 = \mathbf{T}(\psi)$
$\phi_2 = \mathbf{T}(\neg\mathbf{T}(\psi))$
$\psi = \neg(\mathbf{T}(\phi_1) \wedge \mathbf{T}(\phi_2))$

Intuitively, the result should be a similar stable evaluation as in the ordinary Gupta puzzle, provided we can use the Tarski biconditionals without any restrictions. If we assume that Tarski's biconditionals hold, we have the following biconditional: $\phi_2 \leftrightarrow (\neg\mathbf{T}(\psi))$. Then, we are able to reason as in Example 9.1.7. Literally, ϕ_1 and ϕ_2 do not longer contradict each other. According to the following table there is an initial hypothesis h, with $h(\phi_1) = F$, $h(\phi_2) = F$, and $h(\psi) = F$, such that the revision process becomes unstable.

	0	1	2	3	4	5	6	7	···
$\phi_1 = \mathbf{T}(\psi)$	T	F	T	T	F	T	T	F	···
$\phi_2 = \mathbf{T}(\neg\mathbf{T}(\psi))$	F	T	F	F	T	F	F	T	···
$\psi = \neg(\mathbf{T}(\phi_1) \wedge \mathbf{T}(\phi_2))$	F	T	T	F	T	T	F	T	···
$\mathbf{T}(\phi_2)$	F	F	T	F	F	T	F	F	···

The table shows that under certain hypotheses the revision process does not yield a stabilization point. We have to conclude that this is contrary to the intuition and ordinary common sense reasoning.

The considerations show that many of the standard circular sentences (or sets of sentences) can be adequately modeled in the revision theoretic account. There are also examples where revision theories do not provide the intended analysis. Most prominently these examples are the Truth-teller circle and the modified Gupta puzzle. We will reconsider possible ways out of the described problems in Subsection 10.1.2 and Subsection 10.1.3 where modified accounts will be presented which block the unstable evaluation sequences.

In the next sections, we will consider further applications of revision theories. Although the main focus and endeavor of most authors was to apply revision theoretic techniques to a theory of truth there are other applications as well. These applications range from philosophical considerations to quite technical applications in set theory. In the following section, we will begin with some philosophical considerations concerning the nature of truth, i.e. the general assumptions that are necessary to develop the revision theoretic framework.

[12] This example was already mentioned in [GuBe93].

9.2 The Nature of Truth

In [GuBe93], Anil Gupta and Nuel Belnap formulated two central claims concerning the properties of their analysis of the truth predicate for natural languages. Moreover, these two claims can count as the philosophical basis of their whole theory.

- The truth concept is essentially a circular concept, but it is not an inconsistent concept.
- There do exist appropriate semantical frameworks in which circular concepts (as for example the truth concept) make sense and can be modeled appropriately.[13]

A further important assumption of Gupta and Belnap is the consistency of Tarski's biconditionals. This is crucial because a standard argument is that Tarski's biconditionals together with a classical (two-valued) logic, a classical syntax, and a classical semantics yield inconsistencies. This is the crucial upshot of Tarski's famous theorem concerning the undefinability of truth in sufficiently strong languages. Because the above statements are not mathematical statements but similar to philosophical principles, it is clear that one can only give arguments for or against these claims. An ultimate proof of the statements above is not possible.

A standard criticism of the claim that the truth concept is a circular concept is the thesis that not the truth concept causes problems but the language in which a truth concept is introduced. According to this thesis the crucial point is the strength of the language and not the truth concept, because in a language that does not have an elementary coding scheme for the own expressions pathological sentences cannot be built.

Following this criticism shifts the problem to another level. Although there is a certain appeal to this latter claim[14] and paradoxes can also be built with predicates not interpreted as truth predicates, it is questionable whether this criticism is correct for natural language. First, it is true that natural language is strong enough to code paradoxes. Interestingly enough, although there are numerous (natural) predicates and concept in natural language we deal with, there is not one predicate that causes the same perennial problems like the truth predicate. Only the truth predicate is a natural and ordinary concept that has this properties. Other circular phenomena and pathological entities are higher-order concepts of abstractions (like circularity in mathematical contexts, linguistics, or computer science) or result in the attempt to model certain general capacities (as for example the attempt to model common knowledge).

There is a second argument against the thesis that the property of languages causes the problem and not the truth predicate. This has to do with the fact that the truth predicate is a predicate that determines a language (in a certain

[13] Cf. [GuBe93], p.253.

[14] One has to take into account that the hierarchies of formal languages show that strong languages have properties that are usually not wanted. For example, non-decidability or the property to be non-recursive are often consequences of the strength of a particular theory.

sense). It is probably too strong to claim (like Donald Davidson does) that to know the semantics of a natural language is equivalent to the knowledge of the true sentences of a language. Seemingly it is true that the knowledge of the truth predicate of a language implies to know the semantic value of all simple statements. But to know the semantics of a natural language is more than to know the truth values of simple statements. One can say that the knowledge of the truth predicate is a necessary condition for the knowledge of the semantics. Hence, the truth predicate *is* in a certain sense the problematic aspect, because it emphasizes the problem, whereas the claim that the strength of the language is the major problem is rather uninteresting.

Accepting that the truth concept causes the problem instead the strength of the language, comes quite close to accept that truth is a circular concept causing its own problems because of its circular character. Furthermore, this implies that circularity is a property that essentially plays a certain role in the cognitive abilities of human beings.

The claim that the truth predicate is the concept that causes the problems is an intuitive idea. This cannot be said about the second claim of Gupta and Belnap, namely the claim that the truth predicate is not inconsistent. From a naive perspective it seems to be the case that the mere possibility to deduce contradictions from using the truth predicate in natural language implies that the truth concept is an inconsistent concept. Essentially, Gupta and Belnap's program is to show that there is the possibility to model the (circular) truth predicate in a two-valued logic and to remain nevertheless consistent in their analysis. The interpretation of the usage of the truth predicate can be formulated as follows: although there are paradoxes it is not the case that natural language is inconsistent.[15]

The precise formulation of the Gupta-Belnap systems show that there is a semantical theory for circular expressions (and these systems are provably consistent even though they are not decidable) and that the modeling of these expressions is quite often the intuitively correct one. In this respect, the claim of Gupta and Belnap is correct. A more challenging question is whether revision theories in fact model the truth predicate of natural language or rather model an idealized and artificial truth predicate that is not equivalent to the truth predicate in natural language. It is obvious that a claim like "the truth predicate determined by the semantical system $\mathbf{S}^*$ *is* the truth predicate of natural language" cannot be correct. A hint that we are dealing with an infinite number of idealized versions of a truth predicate can be given by the fact that there is an infinite number of semantical systems Gupta and Belnap propose in their work.[16] Additionally, it is simply not true that Gupta-Belnap systems do not have problematic features and solve every possible problem of modeling the

[15] This was probably one of the very basic intuitions when people developed paraconsistent logics.

[16] We only considered the most important semantical systems $\mathbf{S}^*$, $\mathbf{S}^\#$, and $\mathbf{S}_i$. In [GuBe93], there are infinitely many more semantical systems, namely the systems $\mathbf{S}_n$ for $n \in \omega$.

truth predicate.[17] Although it is questionable whether revision theories give a precise and totally correct representation of the truth concept, it is nevertheless true that they give a reliable account for a circular concept. Therefore, the claim that it is in principle possible to develop a consistent semantical theory (in a two-valued logic) for a circular concept can be affirmed.

In total, one can say that Gupta and Belnap proposes a framework in which circular concepts can be modeled and furthermore that the truth concept is a circular one. As we pointed out at the beginning of this section one cannot establish a proof that the truth predicate is a circular concept, but there are good arguments for such a claim. We think that the alternative, namely to lift the pathological behavior of truth to a problem of languages is begging the question and does not solve any problem.

We will consider further applications of Gupta-Belnap systems in the following section. We divide these applications in different topics in order to get a better overview which entities can be modeled using revision theories and which cannot.

9.3 Other Applications

9.3.1 Hypersets Revision Theoretically Defined

Although we will develop the theory of hypersets in Part IV of this work quite explicitly, it is illuminating to mention an interesting application of revision theories in set theory. Using revision theoretic techniques it is possible to construct a model of set theory that contains non-well-founded sets, i.e. sets that contain themselves as elements. Examples for such sets are: $a = \{a\}$ or $b = \{b, c, \emptyset, 4\}$.[18] We do not want to introduce the precise machinery of this theory, in particular, we do not want to show that a model of non-well-founded set theory can be defined using revision theories, but we try to give an idea how the account works principally. A full development of the construction of a non-well-founded universe using Gupta-Belnap systems can be found in [An94b]. The reader is referred to this work for further information.

First, we need to say something concerning basic set theory.[19] We assume that the universe of well-founded sets V is given. Our set theory if based on the standard axioms ZFC. The foundation axiom states that every non-empty set a in the universe contains an element y, such that y is minimal with respect to the non-logical relation $\in$.[20] This axiom prevents the existence of sets of the form $a = \{a\}$ or $b = \{b, c, \emptyset, 4\}$. Any set that includes itself is not allowed

[17] For a discussion of these problematic features compare Chapter 10 and especially Section 10.1.

[18] Clearly, sets like a and b are not in the well-founded universe $\mathcal{V}$, because they contradict the foundation axiom.

[19] An axiomatic introduction and discussion of standard set theory can be found in Chapter 11, especially in the Sections 11.1 and 11.2.

[20] Compare Section 11.1 for a precise formulation of the foundation axiom.

as a respectable set theoretic object. Our task is: We want to know (on an informal level) how we can define revision theoretically a set Ω that satisfies the following equation:

$$\Omega = \{\Omega\}$$

In particular, we would like to use revision theoretic techniques to extend the ordinary universe of well-founded sets with the new set Ω satisfying the above equation. The idea of how this can be achieved is to extend the ordinary relation $\in$ to a relation $\widehat{\in}$, such that $\widehat{\in}$ restricted to well-founded sets is equal to $\in$. Only with respect to the object $\Omega = \{\Omega\}$ the extended elementhood relation $\widehat{\in}$ satisfies $\Omega \mathrel{\widehat{\in}} \Omega$.

We introduce some notions and concepts. Because of our simplified presentation we need only to define the properties of an object $\Omega = \{\Omega\}$ with respect to the extended elementhood relation $\widehat{\in}$. As we will see later in Chapter 11, urelements can be used to define systems of equations. These systems are an appropriate tool for an extension of the universe.[21] Let us assume that an urelement Ω is given. An important point is to introduce an extension of the principle of extensionality of ZFC that can be applied to non-well-founded sets, too. We state the principle of extensionality as follows:

$$(\forall a)(\forall b)(\forall c)(a = b \;\leftrightarrow\; (c \in a \;\leftrightarrow\; c \in b))$$

The following definition gives us an account for an extended principle of extensionality:[22]

Definition 9.3.1 *Define for arbitrary sets of the universe V and the urelement Ω the binary relation $\equiv_R \;\subseteq\; (V \cup \{\Omega\}) \times (V \cup \{\Omega\})$ as follows:*

$$a \equiv_R b \;\;\Leftrightarrow\;\; \forall c\,(\langle c, a\rangle \in R \leftrightarrow \langle c, b\rangle R) \wedge (a = \emptyset \leftrightarrow b = \emptyset)$$

Remark 9.3.1 (i) Notice that the above definition is quite similar to the extensionality principle if we substitute the elementhood relation $\in$ for the relation R in Definition 9.3.1. The relation $\equiv_R$ functions as a modified version of an extensionality principle that can be used also for non-well-founded sets.

(ii) The idea of the whole construction is to use the extension of R as an hypothesis. Step by step, this extension R will be improved by the revision process. R itself is an initial hypothesis (or what is sometimes called a bootstrapper) of the revision process.

[21] For more information concerning urelements compare Chapter 11. Urelements are objects that are themselves non-sets but that can be used in order to build sets.

[22] As a side remark it should be mentioned that this is not a bisimulation relation.

Now we can define a modified elementhood relation $\widehat{\in}$ that has the following behavior. Given a hypothesis R, the following definition gives us a new extension of the modified elementhood relation.

Definition 9.3.2 *We define the jump operation of the modified elementhood relation provided a hypothesis R is given as follows:*

$$\forall a \forall b : a \mathrel{\widehat{\in_R}} b \;\;\Leftrightarrow\;\; (b \neq \Omega \rightarrow \exists c(a \equiv_R c \wedge c \in b)) \;\wedge\; (b = \Omega \rightarrow a \equiv_R b)$$

In Definition 9.3.2, the modified elementhood relation $\widehat{\in_R}$ works on well-founded sets as the ordinary elementhood relation $\in$. Notice further that $\Omega \mathrel{\widehat{\in_R}} \Omega$ holds (using Definition 9.3.1). In a certain sense, this is already the crucial point of the whole development: The jump operation respects the ordinary elementhood relation and extends this relation to the non-well-founded set Ω.

The reason why we do not introduce the complete machinery here and work only with this very special non-well-founded set Ω is based on the fact that in order to develop a more general picture one has to introduce the precise concept for coextensionality of non-well-founded sets, namely a congruence relation that is called bisimulation. This concept will be introduced and discussed in Section 11.4.

Now we are ready to define the revision process that generates step by step (independently of the initial hypothesis) the new extension of the elementhood relation $\widehat{\in}$.

Definition 9.3.3 *Assume $R \subseteq (V \cup \{\Omega\}) \times (V \cup \{\Omega\})$ is an arbitrary relation. The revision sequence defining the extension of $\widehat{\in_\beta}$ (for $\beta \in ORD$) is specified as follows:*

$$a \mathrel{\widehat{\in_0}} b \;\Leftrightarrow\; \langle a, b\rangle \in R$$

$$a \mathrel{\widehat{\in_{\alpha+1}}} b \;\Leftrightarrow\; a \mathrel{\widehat{\in_{\widehat{\in_\alpha}}}} b$$

If λ is a limit ordinal we have to distinguish three cases, dependent on the stability properties of former revisions and similarly to the condition in Definition 8.1.1(ii).

$$a \mathrel{\widehat{\in_\lambda}} b \iff \exists \alpha < \lambda \forall \beta : (\alpha \leq \beta < \lambda \;\rightarrow\; a \mathrel{\widehat{\in_\beta}} b)$$

$$a \mathrel{\widehat{\notin_\lambda}} b \iff \exists \alpha < \lambda \forall \beta : (\alpha \leq \beta < \lambda \;\rightarrow\; a \mathrel{\widehat{\notin_\beta}} b)$$

Otherwise it holds: $\langle a, b\rangle \in R$.

Remark 9.3.2 The specification of the revision process in Definition 9.3.3 is quite abstract, therefore we state some intuitions that are constitutive for it. The extension of the initial hypothesis can be arbitrarily chosen: $R \subseteq (V \cup \{\Omega\}) \times (V \cup \{\Omega\})$. In our case, it is easy to check that every hypothesis would work: every initial hypothesis results in a convergent sequence of hypotheses (provided we work with revision sequences of length ORD).

The successor step revises the extension of the hypothesis according to Definition 9.3.2. This step guarantees that $\langle \Omega, \Omega \rangle$ will be in the extension of $\widehat{\in_R}$ and that the restriction $\widehat{\in_R} \restriction V \times V$ is coextensional to the ordinary elementhood relation $\in$.

Finally, consider the limit stage for a limit ordinal λ. The three cases mirror simply the possibilities that can arise: if an element $\langle a, b \rangle$ is stable in the extension of $\widehat{\in_R}$, then it should remain in the extension. If an element $\langle a, b \rangle$ is stable in the anti-extension this element should remain in the anti-extension. The most interesting case is what we do with unstable elements. Here, the proposed procedure is to use the initial hypothesis R again. This is essentially what was called the Herzberger limit rule above and what was proposed in [He82a] and [He82b]. Notice that the Gupta limit rule would do it as well and also the Belnap limit rule.

In order to get a more intuitive picture of the behavior of non-well-founded sets, we give some more information concerning the properties of Ω. Suppose it holds $\Omega = \{\Omega\}$. Then, obviously the following (infinite) chain of equalities do also hold:

$$\Omega = \{\Omega\} = \{\{\Omega\}\} = \{\{\{\Omega\}\}\} = ...$$

Additionally it holds $\Omega = \{\Omega, \{\Omega\}\}$, because $x = \{y, y\} = \{y\}$. Therefore, we have a variety of sets that are all equal to Ω and that are considered as distinct objects in our revision process. In order to specify Ω as the representative of a collection of objects, one needs to introduce an equivalence class for all these elements. The following definition specifies this equivalence class.

Definition 9.3.4 *Assume Ω satisfies the equation $\Omega = \{\Omega\}$. By transfinite recursion we define the following equivalence classes $[\Omega]^\beta$:*

$$[\Omega]^0 = \{\Omega\}$$

$$[\Omega]^{\alpha+1} = \{x : x \neq \emptyset \wedge x \subseteq [\Omega]^\alpha\} \cup [\Omega]^\alpha$$

$$[\Omega]^\lambda = \bigcup_{\alpha<\lambda} [\Omega]^\alpha$$

Define $[\Omega]$ by the following condition:

$$a \in [\Omega] \;\Leftrightarrow\; \exists \alpha \in ORD : a \in [\Omega]^\alpha$$

Definition 9.3.4 provides an intuitive picture of the properties of the non-well-founded equivalence class $[\Omega]$. Our revision process mirrors the properties of $[\Omega]$ quite nicely, as can be seen by the following fact.

Fact 9.3.5 *For every relation $R \subseteq (V \cup \{\Omega\}) \times (V \cup \{\Omega\})$ and every $a \in [\Omega]$ it holds:*

$$\exists \alpha \in ORD : a \in [\Omega] \Leftrightarrow a \mathrel{\widehat{\in_\alpha}} \Omega$$

Proof: We only sketch the proof of the claim. The interested reader is referred to [An94b] for a more explicit proof.
"$\Rightarrow$" It is necessary to determine the rank of a, rk(a). The claim is proven by (a transfinite) induction on the rank of a. $\alpha = 0$ is trivial. For a successor ordinal $\alpha = \beta + 1$ assume $a \neq \Omega$. Because a is non-empty, there exists $b \in a$, with $b \in [\Omega]$. Using the induction hypothesis we have $b \mathrel{\widehat{\in_\beta}} \Omega$. Because $b \in a$ we have $b \mathrel{\widehat{\in_{\beta+1}}} a$ and therefore $a \mathrel{\widehat{\in_{\beta+1}}} \Omega$. For a limit ordinal α, the claim $a \mathrel{\widehat{\in_\alpha}} \Omega$ follows from the induction hypothesis.

"$\Leftarrow$" Again the proof is an induction on the ordinal α. The proof is similar to the left-to-right direction. q.e.d.

Remark 9.3.3 We summarize the examinations we presented here. The considerations show that it is possible to extend the ordinary elementhood relation $\in$ to a modified elementhood relation $\widehat{\in}$, such that the two relations are equal on well-founded sets and additionally $\widehat{\in}$ captures the properties of a set satisfying the equation $\Omega = \{\Omega\}$. Furthermore, the extension of the binary relation $\widehat{\in}$ can be generated by a revision process in Gupta-Belnap systems. This suffices to give an idea how revision rules can be used in order to construct non-well-founded sets and, in particular, the extension of the elementhood relation $\in$ to $\widehat{\in}$.

The sketched idea how to introduce non-well-founded sets via revision rules is a very special case of a more general idea. Similarly to the case where the universe $\mathcal{V}$ is extended by an equation $\Omega = \{\Omega\}$ one can generalize this idea to arbitrary equations of the form $\Omega = a$ where Ω functions like an indeterminate x in algebra in this context. Although the enlarged framework is slightly more complicated because one needs to introduce a so-called bisimulation relation between sets (or a certain kind of congruence relation as specified in [An94b]), the overall procedure is precisely the same as in the simple case we considered in this section. Interestingly enough, the revision process remains the same in the general case, only the proofs are slightly more complicated.

We finish this subsection with these remarks. A complete development of the applications of revision theories in order to introduce non-well-founded sets can be found in [An94b]. In the next section, we give some more ideas how Gupta-Belnap systems can be used in mathematics.

9.3.2 Other Applications in Mathematics

In this subsection, we mention two other applications of revision theories. Aldo Antonelli[23] showed that revision theories can be used to characterize the arithmetical hierarchy. From an intuitive perspective this should not be very surprising, because the arithmetical hierarchy can essentially be defined inductively and inductive relations are a proper subclass of definable predicates in $\mathbf{S}^*$. Notice that the complexity classes of the arithmetical hierarchy are defined as follows:

$$\begin{aligned}&\forall n \in \mathbb{N} : \forall X \subseteq \mathbb{N} : (X \in \Sigma_0^n \Leftrightarrow \forall x \in \mathbb{N} :\\ &\qquad (x \in X \Leftrightarrow \exists y_1 \forall y_2 ... Q y_n : \phi(y_1, y_2, ... y_n, x)))\end{aligned}$$

$$\begin{aligned}&\forall n \in \mathbb{N} : \forall X \subseteq \mathbb{N} : (X \in \Pi_0^n \Leftrightarrow \forall x \in \mathbb{N} :\\ &\qquad (x \in X \Leftrightarrow \forall y_1 \exists y_2 ... Q y_n : \phi(y_1, y_2, ... y_n, x)))\end{aligned}$$

Dependent on the properties of n, namely whether n is odd or even, we specify $Q = \exists$ for odd n and $Q = \forall$ for even n in the first condition. Similarly we specify $Q = \forall$ for odd n and $Q = \exists$ for even n in the second condition. As can be seen by the characterization of the arithmetical hierarchy, the complexity is at most a Π_1^1 relation. Hence, this relation can be revision theoretically defined.

We state the general idea of such a revision theoretic characterization in the following. Assume we consider a predicate G, using a circular definition with a parameter ϕ:

$$G(x_1, x_2, ..., x_n, \phi) = \ldots \phi \ldots$$

Then, applying a bootstrapper ψ and a revision process that uses the newly calculated extension for a better approximation of G, we can approximate complexity classes of the arithmetical hierarchy. The following definition makes these intuitive ideas precise.

Definition 9.3.6 *Assume $G(x_1, x_2, ..., x_n, \phi)$ is an $n+1$-ary functional. Assume further that $\psi : \mathbb{N}^n \longrightarrow \mathbb{N}$ is total. The revision theoretical definition of the functional G using ψ as initial hypothesis is defined by recursion as follows:*

$$G_0^\psi(x_1, x_2, \ldots, x_n) = \psi(x_1, x_2, \ldots, x_n)$$

$$G_{\alpha+1}^\psi(x_1, x_2, \ldots, x_n) = G(x_1, x_2, \ldots, x_n, G_\alpha^\psi)$$

$$G_\lambda^\psi(x_1, \ldots, x_n) = \begin{cases} z & : (\exists \sigma < \lambda)(\forall \tau)(\sigma < \tau \rightarrow G_\tau^\psi(x_1, \ldots, x_n) = z) \\ \psi(x_1, \ldots, x_n) & : \textit{otherwise} \end{cases}$$

[23] Cf. [An94c].

Remark 9.3.4 (i) The initial hypothesis ψ is assumed to be total. We chose a total function instead of a partial function because it simplifies matters. Whereas complexity considerations using functionals with partial functions as arguments are quite complicated, it is easier to use total functions. Additionally, a total function ψ suffices to get a characterization of the arithmetical hierarchy.

(ii) For the limit step Antonelli (in [An94c]) used the constant limit rule where the initial hypothesis is used in order to assign a value to the unstable elements (Herzberger limit rule).

(iii) Definition 9.3.6 resembles an inductive definition. This supports the claim that there is a certain similarity between inductive definitions and revision theories. The differences between both accounts are based on the various possibilities to choose the initial hypothesis and the fact that the revision process need not to be a monotone process. Precisely the fact that revision processes need not to be monotone is the reason why it is necessary to develop an extension of classical semantics. This extension guarantees that it is possible to evaluate arbitrary formulas in the revision theoretic case.

The upshot in [An94c] is that revision theories can be used to define the arithmetical hierarchy. Interestingly enough, it is not necessary to consider revision sequences of length ORD. It suffices to consider a strongly restricted set of revision sequences. We state the main result in [An94c], namely the characterization of the arithmetical hierarchy using revision sequences of length at most ω^2.

Theorem 9.3.7 *The set of true sentences of first order arithmetic is many-one reducible to* $G^{\psi}_{<\omega^2}$

Proof: Compare [An94c]. q.e.d.

Whereas a characterization of the arithmetical hierarchy can be achieved it is clearly not possible to characterize the analytical hierarchy. The reason for this is that the complexity of revision theoretically definable subsets of ω is bound by Δ^1_3. We cannot go behind this border.

Another application of revision theories is the connection between inductive definitions and revision theories. In the following, we add some additional remarks concerning this context, in particular, we show that fixed point theories can be modeled by Gupta-Belnap systems.

9.3.3 Fixed Points and Revision Theories

In Part II of this work, we examined a fixed point construction in order to generate partially defined truth predicates. Within certain limits these fixed points represent possible truth predicates of natural language. Essentially, Kripke's construction is an inductive definition of (different types of) fixed points. We saw in Section 8.3 that every inductive definition (Π^1_1 definition) can be represented as a circular definition in $\mathbf{S}^*$. Therefore, there exists a possibility to represent Kripke's fixed point approach using revision theoretical means. We will consider this idea more closely in this subsection.

We remind the reader that every inductive definition can be represented and evaluated in $\mathbf{S}^*$as a circular definition. Assume X is an inductively defined subset of ω, such that $A(x, G)$ denotes the definiens of the inductive definition of X. Then, the predicate H as specified below is coextensional with the inductively defined predicate.[24]

$$G(x) \Leftrightarrow \forall x(G(x) \rightarrow A(x,G)) \wedge \exists x \neg(A(x,G) \rightarrow G(x)) \wedge A(x,G)$$

$$\begin{aligned} H(x) \Leftrightarrow {} & [\forall x(G(x) \rightarrow A(x,G)) \wedge \forall x(A(x,G) \rightarrow G(x)) \wedge G(x)] \ \vee \\ & [\neg(\forall x(G(x) \rightarrow A(x,G)) \wedge \forall x(A(x,G) \rightarrow G(x))) \wedge H(x)] \end{aligned}$$

For a moment we restrict our attention to the minimal fixed point in Kripke's construction. Assume the set of functions $\{T, F, N, B\}^{Sent_{L^+}}$ is given (relative to a given language L). The function space can be partially ordered by $\leq_i$ where $f \leq_i g$ holds if and only if for all $d \in Sent_{L^+}$ it holds: $f(d) \leq_i g(d)$ (relative to the interlaced bilattice $\{\langle\{T, F, N, B\}, \leq_I, \leq_T\rangle\}$. The bottom element of this bilattice in the information order is the function $f : \phi \longmapsto N$ for all $\phi \in Sent_{L^+}$. Assume further that a monotone operator $\Gamma : \{T, F, N\}^{Sent_{L^+}} \longrightarrow \{T, F, N, B\}^{Sent_{L^+}}$ is given. A transfinite induction on the ordinals shows that there is $\lambda \in ORD$, such that $\Gamma^\lambda(\bot) = \Gamma^{\lambda+1}(\bot)$, i.e. that $\Gamma^\lambda(\bot)$ is a fixed point. Let us call this fixed point Γ_*. We will construct this minimal fixed point Γ_* revision theoretically.

Definition 9.3.8 *Assume that a language $L^+ = L \cup \{\mathbf{T}\}$ and a classical ground model $\mathfrak{M}$ is given (similar to Definition 8.1.4). Assume further that a monotone operator $\Gamma : \{T, F, N, B\}^{Sent_{L^+}} \longrightarrow \{T, F, N, B\}^{Sent_{L^+}}$ is given. Finally, assume that $\bot$ is the constant function $f : Sent_{L^+} \longrightarrow \{T, F, N, B\} : \phi \longmapsto N$. Our revision process in order to construct the minimal fixed point of Kripke's construction is defined as follows.*

$\mathbf{T}_0 = \bot$
$\mathbf{T}_{\alpha+1} = \Gamma(\mathbf{T}_\alpha)$
$\mathbf{T}_\lambda = \sup\{\mathbf{T}_\beta\}_{\beta<\lambda})$

Remark 9.3.5 (i) It is clear that the resulting revision sequence $\langle\bot, \Gamma(\bot), \Gamma(\Gamma(\bot)), ...\rangle$ is precisely the transfinite inductive process well-known from constructing minimal fixed points.

[24] Compare Lemma 8.3.2.

(ii) Important in the above considerations is the fact that we used a specific initial hypothesis $\perp$ in order to start the revision process. Although there are infinitely many different initial hypotheses in order to get a sequence of hypotheses that finally result in a fixed point, it is not true that every initial hypotheses works for a particular fixed point (for example the minimal fixed point). The reason for this is the fact that there are infinitely many fixed points that can be reached by a revision process if one varies the initial hypotheses.

(iii) Notice that every initial hypothesis can be used in order to generate recurring hypotheses. This is easily seen by the fact that the operator Γ and therefore our revision rule ρ is monotone.

(iv) It should be mentioned that the collection of all recurring hypotheses are not coextensional with the collection of all fixed points of $\Gamma : \{T, F, N, B\}^{Sent_{L^+}} \longrightarrow \{T, F, N, B\}^{Sent_{L^+}}$.

We state an important additional remark concerning the usage of the non-classical ground model in the considerations so far.

Remark 9.3.6 (i) The term *classical ground model* is slightly misleading, because the underlying logic is not a two-valued classical logic but a four-valued non-classical logic. That does not create problems, because models for partial logics are well-known and their behavior and their properties are well-understood. In contrary to the Gupta-Belnap systems $\mathbf{S}^{\#}$ and $\mathbf{S}^{*}$ that are not recursively axiomatizable (compare 8.3.1(i)), classical ground models in four-valued (as well as three-valued) logic are axiomatizable. Furthermore, the compactness theorem, the Robinson theorem, and the Löwenheim-Skolem property are preserved in these models. For an overview of non-classical logics and their properties the reader is referred to [Mu89] or [Kl52].

(ii) The Gupta-Belnap systems are general enough to be applied to other frameworks like non-classical logics. It is important to notice that revision theories are applied to a non-classical framework on a meta-level. Although from a mathematical perspective there are no problems to apply the revision theoretic approach to a four-valued framework, the philosophical perspective of this situation is not as simple. Clearly the development of revision theories was essentially motivated by the attempt to find a framework for pathological sentences where classical logic can be preserved. The idea was to find a theory that is able to model Liar-like sentences but furthermore does not abolish classical logic. Therefore, showing that the fixed point construction of Kripke can be modeled revision theoretically is a mathematical fact, but has no philosophical implications for someone who affirms the philosophical claims of revision theories.

A slightly more complex construction allows us to construct the set of all fixed points using a revision theoretic process. The next definition states this construction precisely.

Definition 9.3.9 *Assume a language L and a classical ground model $\mathfrak{M}$ are given. Assume further that Γ is a monotone operator with the property $\Gamma : \{T, F, N, B\}^{Sent_{L^+}} \longrightarrow \{T, F, N, B\}^{Sent_{L^+}}$. Consider the set $F = \{f \mid f : Sent_{L^+} \longrightarrow \{T, F, N, B\}\}$. The revision process generating all fixed points of the Kripke approach is defined as follows:*

$$\begin{aligned}
&\{\mathbf{T}_0^f\}_{f\in F} = \{f \mid f : Sent_{L^+} \longrightarrow \{T, F, N, B\}\}\\
&\{\mathbf{T}_{\alpha+1}^f\}_{f\in F} = \{\Gamma^\alpha(f) \mid f : Sent_{L^+} \longrightarrow \{T, F, N, B\}\}\\
&\{\mathbf{T}_\lambda^f\}_{f\in F} = \bigcup\{\Gamma^\beta(f) \mid \beta < \lambda \ \wedge \ f : Sent_{L^+} \longrightarrow \{T, F, N, B\}\}
\end{aligned}$$

Although the next fact is quite straightforward to see we state it explicitly. It shows that the collection of all fixed points in a Kripke-style construction can be modeled using revision theoretic techniques.

Fact 9.3.10 *The revision process described in Definition 9.3.9 contains stable extensions for an appropriate ordinal $\lambda \in ORD$. Furthermore, for an appropriate ordinal λ' the set $\{\mathbf{T}_{\lambda'}^f\}_{f\in F}$ restricted to stable extensions is isomorphic to the collection of all fixed points of an ordinary Kripke-style construction.*

Proof: The existence of stable extensions follows from the facts that Γ is monotone and monotone operators do have fixed points (compare [Bar75]). The fact that the set $\{\mathbf{T}_\lambda^f\}_{f\in F}$ contains all fixed points of a Kripke-style construction is obvious. q.e.d.

Remark 9.3.7 (i) The generalization of the simple case of Definition 9.3.8 where one fixed point is revision theoretically defined to the case in Definition 9.3.9 where infinitely many fixed points are considered simultaneously is straightforward. The difference is simply to collect all possible functions mapping $Sent_{L^+}$ into $\{T, F, N, B\}$ and start the revision process from above, i.e. shrinking this set down to the functions that are fixed points (or are unstable). Notice that in a set theoretical representation it is not necessary to introduce indices for the single functions $f : Sent_{L^+} \longrightarrow \{T, F, N, B\}$.

(ii) Notice that the revision process shrinks the total number of elements in the collections of all functions $\{f \mid f : Sent_{L^+} \longrightarrow \{T, F, N, B\}\}$ to the set $\{\mathbf{T}_\lambda^f\}_{f\in F}$ that contains only fixed points. The reason for this phenomenon is based on the fact that many different initial hypotheses result in the same collection of fixed points when applying the revision operator Γ.

(iii) The fact that Kripke's fixed point approach is essentially an inductive account, makes it not very surprising that it is possible to represent his construction revision theoretically. We used as initial hypothesis h the collection of all possible functions mapping $Sent_{L^+}$ into $\{T, F, N, B\}$. This choice is important, because not every subset of h guarantees that all fixed points can be reached. Furthermore, notice that Gupta-Belnap systems (in particular, the systems $\mathbf{S}^*$ and $\mathbf{S}^\#$ as well as the finite systems $\mathbf{S}_n$) were developed even to model non-monotonic revision rules, that do not necessarily converge in fixed points. This is a further difference between the two accounts mentioned so far.

In the final section of this chapter, we will summarize the different applications of Gupta-Belnap systems we mentioned so far and make some final remarks concerning possibilities to model propositions revision theoretically. Similar to most presentations of this chapter we do this informally.

9.4 Some Final Remarks

In the above sections, we examined a variety of applications of Gupta-Belnap systems. We mentioned some classical applications like the development of a theory of truth and the endeavor to give a precise and intuitively correct analysis of pathological sentences. Furthermore, we stated some implications for a theory of truth and we mentioned (although not completely and explicitly developed) some applications of Gupta-Belnap systems in a more technical context like its relation to non-well-founded set theory, the arithmetical hierarchy, and Kripke's fixed point approach.

It is important to emphasize that we did not mention applications concerning knowledge representation, common sense reasoning (except applications like the modeling of the Gupta puzzle), or applications in the theory of belief revisions. The reason for this is that these applications of revision theories are not spelled out yet. The researchers working on Gupta-Belnap systems usually applied this framework to truth theories and inferred consequences from this modeling for a philosophical theory of truth. Additionally, the quite general assumption of most authors working in this field, namely that the bearers of truth are sentences (and not propositions), makes a treatment of phenomena where the context plays an important role as well as examples where there are more abstract entities included (like beliefs, attitudes, intentions etc.) quite complicated. We should add some comments concerning this situation here.

An application of revision theories to common sense reasoning or belief revisions makes it necessary to enlarge the account of revision theories from the evaluation of sentences to an evaluation of propositions (perhaps even more sophisticated entities). The perennial problem is that a generally accepted theory of propositions does not exist. The authors developing theories for propositions struggle with many different problems arising from the fact that propositions are abstract entities that are highly dependent on the context.[25] How can this be addressed by revision theorists? From

[25] In Section 10.2, we will consider the following two questions: which entity is the bearer of truth and what are the alternatives for revision theories in order to treat not only sentences but also propositions.

Subsection 9.3.1 we know that in principal we can define non-well-founded sets using appropriate revision rules. We will see later (compare Part IV) that non-well-founded sets are an alternative framework to model circular phenomena. As we will see non-well-founded sets are particularly a good framework to model circular propositions.[26] Using this fact should make it possible finding a way to introduce circular propositions with revision theoretic techniques. The idea is simply to remodel the constructions of other theories (like the constructions in situation theory which is clearly only an example).

A possible plan for the development of a theory of belief revisions and knowledge representation in a revision theoretic framework can be given by the following list:

- Define the universe of non-well-founded sets using revision theoretic techniques as developed in subsection 9.3.1.
- Specify particular propositions that are needed in order to model the intended circular phenomena.
- Test the behavior of these propositions in the (interesting) semantical systems $\mathbf{S}^*$ or $\mathbf{S}^\#$.

A different strategy would be the attempt to find a modeling of circular phenomena directly without the detour using non-well-founded set theory. There are good reasons for preferring such a strategy in the case that this second alternative would represent the phenomena in question at least as well as the account using non-well-founded sets. First, avoiding the modification of the underlying set theory guarantees a classical set theoretic framework. Second, classical ZFC set theory is well-developed and well-understood by nearly everybody. This is not true for the theory of non-well-founded sets. Additionally, in a classical set theory only standard ontological problems concerning the status of sets arise, but no deeper questions concerning a justification of the enormous enlargement of the universe needs to be answered. A natural question is the following one: How can this second alternative be achieved?

It is necessary to substitute propositions for sentences. If we assume that this is (somehow) possible, first, we show that the representation of a circular predicate (situation) s in situation theory yields circular sets. This is precisely what we do not want, because this results in the unsatisfactory necessity to introduce non-well-founded sets. In order to see this, assume a complex relation (situation) is given according to the following equation.[27]

$$s = s' \wedge \phi \wedge \langle know, a, s\rangle$$

We can represent the extension of s by the following generalized (and circular) definition:

[26] We will develop this idea in Chapter 16.

[27] The expression $\langle know, a, s\rangle$ is the formal representation of the information that a knows situation s.

$$x \in s \ \Leftrightarrow \ x \in s' \vee x \in \{\phi, \langle know, a, s\rangle\}$$

Even if situation s' is given, we are facing a problem because on a classical set theoretical level, there is no set s that satisfies the condition $\langle know, a, s\rangle \in s$. It seems to be the case that non-well-founded sets are necessary.

A possibility to solve this problem is to use some results developed in [Bar90] where it is shown that the modeling of common ground using inductive techniques is equivalent to the circular approach under the assumption that there are non-well-founded sets. Inductive techniques can be represented in revision theories. In this inductive approach of analyzing common ground (in our example: two persons a and b have common ground that a proposition ϕ holds), we want to generate a set s that has the following properties (i) - (iii):

(i) $s \subseteq \{\phi, \langle know, a, \phi\rangle, \langle know, b, \phi\rangle\}$
(ii) $s' = \langle know, a, x\rangle \in s \ \rightarrow \ (\langle know, a, s'\rangle \in s \ \wedge \ \langle know, b, s'\rangle \in s)$
(iii) $s' = \langle know, b, x\rangle \in s \ \rightarrow \ (\langle know, a, s'\rangle \in s \ \wedge \ \langle know, b, s'\rangle \in s)$

It is obvious that such a set can be revision theoretically defined, because it is inductively defined. Using Lemma 8.3.2 the existence of a revision theoretic definition follows immediately. That means that a revision theoretical modeling of the common ground phenomenon is in principal possible because of the results provided in [Bar90]. The problem of how to introduce propositions in a revision theoretic context remains to be unsolved.

We finish this section with these rather vague remarks concerning a possible treatment of other phenomena of circularity in Gupta-Belnap systems. A closer look concerning the problem of common ground and, in particular, the differences between private knowledge and common ground will be presented in Chapter 16. In this chapter, more remarks concerning the relation between revision theories and the treatment using non-well-founded sets will be added.

9.5 History

Applications of revision theories are primarily based on the truth concept. The problem of truth was the starting point of generalized definition theory as developed in Chapter 8. As references we mentioned [Gu82, He82a, He82b], and [Be82] as the historically first examinations of this account. Besides technical details the monograph [GuBe93] contains also a lot of applications and many deep philosophical considerations. Further work concerning applications of revision theories to the truth concept can be found in [Ya93, Ch93] and [Ch96]. In [Bre92], revision theories are compared with other accounts of theories of truth. The revision theoretic construction of a non-well-founded universe was first shown in [An94b]. The definition of the arithmetical hierarchy by revision sequences was developed in [An94c]. The considerations concerning the representation of Kripke's fixed point approach revision theoretically is (as far as the author knows) new here.

Chapter 10

A Discussion of Gupta-Belnap Systems

In the last three Chapters, we presented the basic ideas of the strategy to strengthen first-order theories using inductive definitions, coinductive definitions and the general framework of Gupta-Belnap systems together with the formal representation of the underlying theory and the associated properties and features. Furthermore, we considered a variety of applications of GB-systems ranging from philosophical to mathematical applications. We postponed a critical discussion of revision theories in these chapters. Some remarks concerning this point can be found in the present chapter. Important for the examinations in this chapter will be the discussion of apparent problems of Gupta-Belnap systems (like the Truth-teller circle or the strengthened Liar sentence). Furthermore, we will examine questions of a more philosophical nature, namely questions concerning the ontological assumptions of revision theories (like the usage of class-like objects or unrestricted quantification over arbitrary models) and the problem whether it is possible to expand Gupta-Belnap systems to a theory that includes also a theory of propositions. We will begin with the some problems occurring in GB-systems concerning a theory of truth.

10.1 Problems of Gupta-Belnap Systems

In this section, we mention some problems of Gupta-Belnap systems that arise from applications of the framework to a theory of truth for natural languages. We will stress two important sources of criticism: first, we will discuss the Strengthened Liar sentence and its consequences to the revision theoretic approach and second, we will reconsider Examples 9.1.5 and 9.1.8, namely two problems we left open in Chapter 9. Finally, we will consider some further problems that were mentioned by some authors in the literature.

10.1.1 The Strengthened Liar Sentence

It is well-known that for most theories of truth the strengthened Liar sentence causes problems.[1] Sometimes it seems to be the case that many authors use the strengthened Liar as a general argument to reject every approach towards a truth theory. A reason for this is the observation that, although modern theories of truth (usually) try to avoid different levels of languages and types of truth predicates, it seems to be the case that reasoning about the strengthened Liar sentence requires such a hierarchy of levels. This was the result we found in Section 6.1 Example (3). What happens to the strengthened Liar sentence in the revision theoretic account? Can we deduce a contradiction from the strengthened Liar also in revision theories? Before we will have a closer look at this question let us consider the ordinary Liar sentence (1) for a moment.

(1) This sentence is false.

The revision theoretic analysis yields the result that (1) is unstable for every possible initial hypothesis (this is trivial to check). Therefore, the classification of (1) is straightforward: (1) is an unstable sentence, hence (1) is essentially paradoxical.

Now, we consider the ordinary strengthened Liar sentence (2) (i.e. the one in Section 6.1 Example (3)):

(2) This sentence is not true.

As we saw in Section 6.1, sentence (2) causes problems in fixed point approaches. A similar problem arises in revision theories: Because we are working in a two-valued logic the behavior of (2) in an evaluation sequence is similar to the behavior of the ordinary Liar sentence (1): (2) is unstable. And precisely this is, according to some authors, questionable. We introduce a special form of the strengthened Liar sentence in order to reconsider this problem more closely. Before we can begin our consideration of this special form of the strengthened Liar sentence below, we need to introduce the concept of categoricalness (of sentences). This new concept, introduced in [GuBe93], allows us to formulate a strengthened Liar sentence that in fact seems to cause problems for revision theories.

Definition 10.1.1 *Assume L is a given language. A sentence ϕ is called categorical if and only if ϕ is either stably T or stably F in all evaluation sequences, hence ϕ is either uniformly T, or uniformly F.*

It was claimed by different authors that the following version (3) of the strengthened Liar sentence causes problems for revision theories:[2] Notice that this version uses two higher-order concepts, one is the concept categoricalness

[1] We saw in Section 6.1 Example (3) that the strengthened Liar sentence is a problem for fixed point approaches in general.

[2] Cf. [GuBe93]. The criticism was mentioned by [Ca86] and [Pr87].

and the other one is the truth concept.

(3) Either this sentence is not categorical or this sentence is not true.

The criticism along these lines can be summarized as follows: assume (3) is categorical, then the first disjunct is false and the whole sentence behaves like the Liar sentence. Therefore, (3) cannot be categorical, because (3) is neither stably true nor stably false. On the other hand, if (3) is not categorical, then the first disjunct is true, hence the whole sentence is stably true which means that (3) is categorical. Together we get: (3) is categorical if and only if (3) is not categorical. Hence, we have a contradiction.

How can revision theories deal with this problem? There are several points that can be mentioned in order to show that this kind of strengthened Liar sentence is a relatively weak argument against revision theories. First, it is clear that without introducing an additional predicate that determines the extension of the predicate *categorical in L* we cannot evaluate (3) using Gupta-Belnap systems in any reasonable way. Therefore, a theory that tries to give an account for truth cannot necessarily solve a problem for which it was not developed, namely to model a theory of categoricalness. As a consequence of this observation, we see that there is not only the truth concept that can cause a pathological behavior, but there are also other concepts that have a similar effect.[3]

Second, the above reasoning that yields a paradoxical behavior of (3) does not take into account that the concept of being *categorical in L* is at least as problematic as the truth concept. As in the case of Liar sentences, if one uses classical logic and a classical semantics, paradoxes can be deduced from *categoricalness in L*. Precisely this was done as part of the reasoning above that finally results in a contradiction. What is needed is a similar revision theoretic treatment of this concept as for the truth predicate. That means we have to introduce a further predicate Cat that must be evaluated in revision sequences using initial hypotheses. The principles that govern the behavior of Cat have to refer to the properties of revision theoretic evaluation sequences of sentences. If L is strong enough to code a sentence like (4), it is easy to see that Cat is a circular concept that results in paradoxes under certain circumstances.

(4) $\phi = \neg Cat(\phi)$

An easy calculation of the extension of Cat shows that (4) is in fact paradoxical.[4] Therefore, the reasoning that results in a contradiction of (3) bears a problem because it does not take into account that Cat itself is a circular concept generating the same problems like the truth concept .

[3] Clearly, there is a connection between the truth concept and the concept to be *categorical in L* in revision theories, because the second concept is defined using crucially the truth-values in revision sequences.

[4] Notice that in this calculation one has to determine the extension of Cat, not the truth value of (4).

At this point, a further remark is useful: in our presentation of Gupta-Belnap systems, we introduced circular definitions as a generalization of classical mathematical definitions except that there is the additional property that the definiendum can occur in the definiens of the definition. The mathematical environment clarifies what is meant by an extension of ordinary definitions. In applications of revision theories to natural language, more can be said concerning the form of the equivalence relation between the definiens and the definiendum. Standardly, in mathematics, the relation between definiens and definiendum is assumed to be a material biconditional. Precisely this fact causes problems when we work in the generalized framework where circular definitions are allowed.[5] That is the reason why Gupta and Belnap (in [GuBe93]) interpret the biconditional in circular definitions as definitional biconditionals. With that interpretation of a definition, the Liar argument which yields inconsistencies is invalid. An example will make this clear. Consider (5) where the symbol $\Longleftrightarrow$ is interpreted as a material biconditional.

(5) This sentence is true $\Longleftrightarrow$ This sentence is not true

With respect to revision theories, there is also the possibility to interpret the biconditional as a definitional biconditional as in (6) where $\Longleftrightarrow_{def}$ is interpreted as a definitional biconditional:

(6) This sentence is true $\Longleftrightarrow_{def}$ This sentence is not true.

Under the first reading, the assumptions that Tarski's biconditionals are consistent with the truth predicate in natural language is inconsistent because we can deduce contradictions. But taken Tarski's biconditionals as definitional equivalences, there is no inconsistency any longer because the Liar sentence is no longer paradoxical but merely false. An account like revision theories shows that under this interpretation it is possible to develop a consistent theory for circular concepts. For Gupta and Belnap, the pathological behavior of the Liar sentence (as well as other pathological examples) does not mean that this sentence needs to be interpreted as neither true nor false because even this move does not prevent the impossibility of paradoxes (as we saw in Section 6.1). Then, the pathological behavior of the Liar sentence is dependent on the behavior of the evaluation sequences, but no longer on a truth value of that sentence. Furthermore, because of the fact that the truth concept is a circular concept, the concept *categorical in L* is necessarily also a circular concept.

We end this subsection with these remarks concerning the strengthened Liar sentence. In the next subsection, we will examine possible solutions of the problems that arose from applications of revision theories to pathological sentences like the Truth-teller circle and the modified Gupta puzzle.

[5] That is one reason that circular definitions are not allowed in mathematics.

10.1.2 The Truth-teller Circle

We saw in Section 9.1 that there are certain natural language examples that cannot be appropriately modeled in the standard account of revision theories we developed in Chapter 8. The classical example for the inappropriateness of the standard theory is Example 9.1.5: the Truth-teller circle. Although the Truth-teller circle is intuitively a TF-capricious example a revision theoretic analysis evaluates this example as TFN-capricious. The problem is that some initial hypotheses result in a paradoxical behavior. How can this problem be fixed?

Historically, two proposals were developed to fix this problem: the first solution is due to Yaqūb[6] and the second solution was proposed by Chapuis.[7] We only consider Chapuis' solution, because Chapuis showed in [Ch96] that Yaqūb's proposal has significant deficits itself: one can construct the same problems in Yaqūb's solution as in the original theory (using some slightly more complex examples). The idea of Chapuis' solution is to change the limit rule of evaluation sequences. We do not allow arbitrary choices at the limit stage of an evaluation sequence of unstable elements any longer, but restrict the considerations to so-called *fully-varied* evaluation sequences. In order to make the idea of a fully-varied evaluation sequence precise, we need to introduce the concept of the coherence of a hypothesis with an evaluation sequence. This concept goes back to [GuBe93].

Definition 10.1.2 *A hypothesis $h \in X^D$ coheres with an evaluation sequence S if and only if for all $d \in D$ and for all $x \in X$ it holds: if d is stably x in S, then $h(d) = x$.*

Coherence of a hypothesis specifies the idea of a certain restriction of hypotheses (relative to a given revision rule), such that it holds: if an element $d \in D$ is stable in S, then all hypotheses h that cohere with S must assign the same extension to d.

Now we give the definition of a fully varied sequence. This concept was informally introduced in [GuBe93]. Here, we adopt the definition of Chapuis in [Ch96].

Definition 10.1.3 *Assume S is an evaluation sequence. We call S a fully-varied sequence if and only if all hypotheses cohering with S are cofinal in S.*

The idea of a fully varied evaluation sequence can be described as follows. If there are stable elements in the revision sequence, then there are hypotheses $h \in X^D$ that are cofinal in S. Although it seems to be the case that this can solve the problem caused by the Truth-teller circle, the proposed definition does not work because there is no a priori condition that determines whether an element is stable or not. Consider again the Truth-teller circle: it is clear that there are revision sequences that make the sentences of the Truth-teller circle stably true (or stably false), but there is no reason why the evaluation sequence

[6] Cf. [Ya93].

[7] Cf. [Ch96]. Although not explicitly, even in [GuBe93], this solution was already proposed.

that assigns an unstable extension should be dismissed. It is vacuously true that every hypothesis that coheres with S (i.e. all hypotheses), is cofinal in S (under the assumption that occurring and unstable hypotheses are repeated over and over again). We need to add some more restrictions.

The problem in the above considerations lies in the fact that all properties of hypotheses are defined relative to a specific evaluation sequence. This special evaluation sequence cannot have properties determining every possible evaluation sequence, in which certain elements can be stable. But this is precisely, what we need. We want an restriction that forces the evaluation to respect stable elements in other revision sequences. What is necessary to introduce as a restriction is some higher-order object that suffices to refer to more than one evaluation sequence. In order to make this idea precise, we use the revision rule $\rho : X^D \longrightarrow X^D$ that determines every possible revision sequence. The following definition correlates coherence of a hypothesis to a restriction that is satisfied by every revision sequence.

Definition 10.1.4 *Assume $\rho : X^D \longrightarrow X^D$ is a revision rule. A hypothesis h coheres with ρ, if and only if for all $d \in D$ and for all $x \in X$ it holds: if there is at least one revision sequence S', such that d is stably x in S', then for every cofinal hypothesis in S, $h(d)$ is stable.*

The idea of this definition is to consider evaluations where elements of D become stable if they can be evaluated as stable in all revision sequences. The next definition clarifies what we mean by the concept that a revision sequence S is varied. This concept will finally be used in order to fix the problem of the Truth-teller circle.

Definition 10.1.5 *(i) Assume $\rho : X^D \longrightarrow X^D$ is a revision rule. We call a revision sequence S varied if and only if all cofinal hypotheses of S cohere with ρ.*

(ii) An evaluation sequence is called appropriate for a theory of truth if and only if all evaluation sequences are varied.

Remark 10.1.1 It is quite obvious that the requirement that all evaluation sequences must be varied suffices to guarantee that the problems described in Example 9.1.5 do not occur. The intuitively incorrect analysis that in a certain revision sequence the Truth-teller circle can be unstable was caused by the choice ϕ *is true* and ψ *is false* (or vice versa) in the initial hypothesis and the same choice (or the reverse choice) at limit stages. Precisely these choices of S are no longer possible in appropriate evaluation sequences, because all cofinal hypotheses in S do not cohere with ρ.

Although the described modification solves the problem of the Truth-teller circle, it implies that there is no evaluation of a sentence (or set of sentences) where we get a TFN-capricious analysis (as well as other analyses like the TN-capricious ones). This is stated by the next fact.

Fact 10.1.6 *Assume $\rho : X^D \longrightarrow X^D$ is a revision rule. If every evaluation sequence is varied, there are no examples for TFN-capricious, TN-capricious, and FN-capricious phenomena.*

Proof: Without loss of generality, assume there was an example ϕ for a TFN-capricious sentence. Then, there is a varied evaluation sequence S, such that all reflexive hypotheses are unstable. Therefore, all cofinal hypotheses do not cohere with ρ. Hence, there is no evaluation sequence S', such that ϕ is stable in S'. This is a contradiction to the assumption that ϕ is TFN-capricious. The same argument can be used in order to show that there are no TN- nor FN-capricious phenomena. q.e.d.

Remark 10.1.2 (i) Fact 10.1.6 shows that the proposed solution limits the possible analyses of examples in a very important respect. Whereas one advantage of revision theories in comparison with fixed point approaches is that revision theories allow a finer distinction of the pathological behavior of certain problematic sentences, precisely this advantage is lost by fixing the problem using varied evaluation sequences. The philosophical question is whether we need the finer distinction between pathological sentences at all. A decision of this question is dependent on the particular application and can vary from case to case.

(ii) From a complexity theoretic perspective the complexity of the systems increase by requiring that all evaluation sequences are varied. In recent work by Philip Welch, it is shown that membership to the categorical truth set becomes a Π^1_3 concept, i.e. this relation is even more complex than the systems $\mathbf{S}^*$ and $\mathbf{S}^\#$.[8]

With respect to an intended solution to find a theory of truth that deals in an appropriate way with pathological sentences, the question whether it should be possible to find examples that are sometimes stably true but sometimes unstable, is difficult to answer. In one respect, the classical pathological examples are examples for biconsistent phenomena (TF-capricious), unstable phenomena (paradoxical phenomena), and stable sentences. A good and undisputed example for a TN-capricious sentence is hard to construct. Consider, for example, the following sentence ψ where λ is the Liar sentence and ϕ is the Truth-teller.

(7) $\psi = \mathbf{T}(\phi) \vee \lambda$

Perhaps (7) is an example that can naively be considered as TN-capricious. It is not easy to argue for such an interpretation, but it goes along the following lines. The Truth-teller sentence is a TF-capricious example and the Liar

[8] Cf. [LöWe∞] for further information.

sentence is unstable. The disjunction of the two sentences could therefore be interpreted as a sentence that is either stably true or unstable (in the case that the Truth-teller is stably true and the Liar sentence is unstable). An easy calculation shows that ϕ is stably true in every evaluation sequence, although (7) includes the Liar sentence. The problem with the above interpretation of (7) is that we interpret the disjunction in a misleading way: we cannot hope that a disjunction combines two disjuncts with a fixed meaning. The disjunction is a connective that respects the relation between the disjuncts: notice that a tautology of the form $p \vee \neg p$ is only a tautology, because the truth value of the first disjunct is true precisely when the second disjunct is false and vice versa. It is not true that the first disjunct is valid or the second disjunct is valid. We will see this misleading interpretation of logical connectives again in Subsection 10.1.4.

In the next subsection, we will reconsider the modified Gupta puzzle again and mention the solution that was proposed by Chapuis.

10.1.3 The Modified Gupta Puzzle

The second problem we mentioned in Section 9.1 was the modified Gupta-paradox (compare Example 9.1.8). According to our intuitive reasoning concerning that example, we want ϕ_1 to be true, ϕ_2 to be false, and ψ to be true (because ψ expresses a logical truth provided Tarski's biconditionals hold). Intuitively, these sentences should have stable truth values. But we found that there is an initial hypothesis that makes the ω revision sequence unstable. If we do not modify the possible choices of the hypotheses of the unstable elements in the limit steps, then we have the similar problem as in Subsection 10.1.2, namely an example for a TN-capricious set of sentences, although intuitively the set of sentences should be stable.

Chapuis' proposal for a solution of this problem is similar to his development of the modified evaluation sequences concerning the Truth-teller circle. We simply require that every evaluation sequence must be a varied revision sequence. That means that the unstable behavior of a revision sequence S as described in Example 9.1.8 cannot occur in an evaluation sequence, because such a sequence is not varied. This solves the problem.

Whereas one can imagine that it is possible to argue against the TF-capriciousness of Example 9.1.5 (at least this seems to be conceivable, because after all the Truth-teller circle is a pathological example), it is without any doubt clear that Example 9.1.8 should be interpreted as a categorical example. Precisely this is achieved by the usage of varied sequences. The question whether the consequence, namely that TFN-capricious phenomena (as well as TN-capricious and FN-capricious ones) are totally banned by this analysis, is acceptable or not remains still open.

10.1.4 Further Problems

Fixed point theories have the advantage to provide a framework in which a truth predicate can be defined in the object language. The price we have to pay is the increase of the number of possible candidates for such a truth predicate (in general there are infinitely many). The possibility to define truth predicates in the object language has as a consequence that we need to determine which fixed point shows the intuitively right properties. At first sight, revision theories do not cause a similar problem because relative to a given semantical system and a given revision rule (and perhaps relative to a certain restriction of evaluation sequences like the ones described in Subsection 10.1.2 and in Subsection 10.1.3), it is clear that also in revision theories we are forced to make a choice: The properties of the truth predicate are dependent on the semantical system we are working in. Additionally, dependent on the particular application, we can modify the restrictions on evaluation sequences in order to prevent counterintuitive analyses. This was the strategy to solve the Truth-teller circle and the modified Gupta puzzle. Because of these alternatives we have - quite similarly to the presented fixed point theories - infinitely many choices between the $\mathbf{S}_n$ systems, the stronger $\mathbf{S}^*$ and $\mathbf{S}^\#$ systems, and special additional features that restrict more or less possible evaluation sequences. So far, Gupta-Belnap systems do have a very similar property as fixed point theories we examined in Part II. Dependent on the particular application we are interested in we are forced to adjust the theory appropriately.

Similarly to the fixed point theories the following question arises: what are sufficient and necessary conditions in order to choose a particular semantical system (or a particular form of an evaluation sequence) in an application. Arguments like the ones stated in Subsection 10.1.2 and Subsection 10.1.3 can help to choose an appropriate form of evaluation sequences. Concerning the choice of the appropriate semantical system, it is clear that for many applications even finite systems suffice. Only for applications where higher-order concepts are needed (as for example in many applications in the field of mathematics) we need to refer to the systems $\mathbf{S}^*$, $\mathbf{S}^\#$, and $\mathbf{S}_n$. Only these systems are strong enough to define second-order concepts.

In a certain sense, revision theories are more flexible in choosing the appropriate systems. Specifying a particular semantical system for a certain application is easier than specifying a particular fixed point. The reason for this is the fact that only some fixed points are real alternatives for a choice (for example the minimal fixed point, the maximal intrinsic fixed point, or one of the maximal fixed points), whereas for many fixed points it is hard to specify which kind of properties they have. In contrast, revision theories can be strengthened or weakened appropriately without changing the general idea behind the modeling.

It is important to notice that revision theories as developed in this part preserve (classical) logical tautologies. As a consequence of this fact a sentence like (8)(i) is stably true in all semantical systems. The following example 8(ii) is the formal representation of (8)(i).

(8)(i) The Liar sentence is true or the Liar sentence is not true.
(ii) $\mathbf{T}(\lambda) \vee \neg\mathbf{T}(\lambda)$

The fact that revision theories are able to preserve logical tautologies was one of the main motivations of the development of the theory. A possible criticism is that (8)(i) is valid although neither of the two disjuncts is true. Yablo, in [Ya85], argues along the line that the validity of (8)(i) is counterintuitive. Furthermore, he claims that this makes the meaning of disjunction mysterious. As Gupta and Belnap in [GuBe93] point out, it is very obscure to criticize the preservation of logical tautologies because the absence of logical tautologies was also criticized in fixed point theories. Moreover, (8)(i) is counterintuitive if one interprets the disjunction in the following way: either the first disjunct is valid or the second disjunct is valid. This is definitely wrong in an ordinary usage of logical disjunction.[9] It seems to be the case that the intuitions of different authors vary. Assuming the classical understanding of logical disjunction should guarantee that Yablo's criticism is on a very weak basis.

The next section deals with the ontological assumptions revision theories need to adopt. In particular, we will consider the impact of class-like objects in revision theories concerning their philosophical implications.

10.2 Ontological Assumptions

In this section, we will consider necessary ontological assumptions for a development of revision theories. The advantage of revision theories in comparison with fixed point approaches is the fact that a classical two-valued logic can be preserved. Furthermore, there is no syntactic restriction blocking pathological sentences like in Tarski's approach. In order to prevent inconsistencies in Gupta-Belnap systems, it is necessary to introduce a semantics that is an extension of classical model theory. This semantics is needed for the evaluation of sequences of hypotheses of length ORD in the case that we work in $\mathbf{S}^{\#}$ or $\mathbf{S}^{*}$ (which are interesting systems for many applications).

We consider two aspects of ontological assumptions. First, we will examine whether sequences of hypotheses of length ORD are needed and second, we will examine to a certain extent the status of higher-order concepts that play a role in Gupta-Belnap systems. Finally, some remarks concerning the sentence vs. proposition distinction concerning the bearers of truth are added.

[9] The criticism has a certain similarity with the criticism of intuitionists concerning classical logic: the principle of the excluded middle is classically valid even if we neither can prove the first disjunct nor the second disjunct. This is not acceptable for an intuitionist. Clearly, this type of criticism cannot be rejected by the proposed analysis.

10.2.1 Sequences of Hypotheses

In our considerations of the semantical systems $\mathbf{S}^*$ and $\mathbf{S}^\#$, sequences of hypotheses have transfinite length, in many cases they are of length ORD. That means that theses sequences are no longer elements of ZFC but each revision sequence is a proper class. This is not a very satisfying result if we take a perspective that considers ontological assumptions. Fortunately, McGee's Theorem (compare 8.2.2) guarantees that every reflexive hypothesis relative to a given revision rule $\rho : X^D \longrightarrow X^D$ can be transformed into an α-reflexive hypothesis where $\alpha \leq \max\{|D|, |X|, \aleph_0\}$, i.e. where α is bound by the size of the model. That means that restricting our attention to countable domains D and countable sets X it suffices to consider simply countable sequences of hypotheses.

McGee's Theorem shows that for a practical application of Gupta-Belnap systems, it is most times sufficient to examine countable revision sequences. Considerations of revision sequences that have length ORD can be avoided in most cases. Even reflexive hypotheses of infinite but countable length are not often needed. Consider again the behavior of the Liar circle 9.1.4 of length 2. We saw in Example 9.1.4 that the Liar circle is unstable for every initial hypothesis. Because the occurring reflexive hypotheses have finite length it is sufficient to consider finite systems. Consider the definition of the much weaker Gupta-Belnap systems $\mathbf{S}_n$.

Definition 10.2.1 *(i) Assume L is a given language, $\mathfrak{M}$ is a given classical model, and let $\rho : X^D \longrightarrow X^D$ be a revision rule. A hypothesis $h \in X^D$ is called n-reflexive iff $\rho^n(h) = h$.*

(ii) Assume L is a given language, $\mathfrak{M}$ is a classical model, and $\phi \in L^+ = L \cup \{\mathbf{T}\}$. ϕ is valid in $\mathfrak{M}$ in $\mathbf{S}_n$ if there exists a natural number $p \in \mathbb{N}$ for all n-reflexive hypotheses h, such that ϕ is valid in $\mathfrak{M} + h$. We denote the relation ϕ is valid in $\mathfrak{M}$ in $\mathbf{S}_n$ by the expression $\mathfrak{M} \models_n \phi$.

Given Definition 10.2.1, we can test the behavior of the Liar circle of length 2 in $\mathbf{S}_n$. This is easy to do and we find precisely the same result as in the general case where we consider (arbitrary) reflexive hypotheses. For the practical application of revision theories concerning the classical Liar-like sentences in natural languages, this shows that we do not need the ontological problematic entities that have transfinite length. In other words, we can model the Liar sentence with finite means.

Clearly, one can construct examples that cannot be appropriately represented and analyzed in the finite systems $\mathbf{S}_n$.[10] For example, what happens with (Liar-like) circles of infinite length? Intuitively, here the situation is different, because sometimes we cannot find appropriate reflexive hypotheses of

[10] We saw already examples in Section 8.2.

finite length. The following example examines the behavior of such an infinite circle of length ω.

Example 10.2.1 Consider the following Liar-like circle:

$$\begin{array}{rcl}
\phi_1 & = & \mathbf{T}(\phi_\omega) \\
\phi_2 & = & \mathbf{T}(\phi_1) \\
\phi_3 & = & \mathbf{T}(\phi_2) \\
\vdots & \vdots & \vdots \quad \vdots \\
\phi_\omega & = & \neg\mathbf{T}(\phi_1)
\end{array}$$

Easy considerations show that this circle is unstable in every revision sequence. The reflexive hypotheses have length ω. This implies that the above transfinite circle shows an unstable behavior in $\mathbf{S}^*$ and $\mathbf{S}^\#$, hence the circle shows a Liar-like behavior. Moreover, because there are no n-reflexive hypotheses (for $n \in \mathbb{N}$), the evaluation of this infinite circle yields the result that the circle is simply false: The evaluation takes place simply in the given ground model $\mathfrak{M}$ and in $\mathfrak{M}$, Liar-like circles cannot be evaluated as true. A more refined examination of this kind of pathological behavior (i.e. unstable behavior with reflexive hypotheses of length ω) is not possible in finite systems $\mathbf{S}_n$. Conclude: the usage of the infinite systems $\mathbf{S}^\#$ or $\mathbf{S}^*$ are appropriate here.

The last example shows quite clearly that sometimes there is the need for systems stronger than the 'finite' Gupta-Belnap systems $\mathbf{S}_n$. Finite sequences of hypotheses cannot capture all the interesting features of certain examples. Additionally, the proposed solutions for the modified Gupta puzzle (compare Example 9.1.8) in Subsection 10.1.3, and the possible ways to fix the counter-intuitive analysis of this example, leads us to the concept of varied sequences of hypotheses. Here, again it seems reasonable to assume that we need systems stronger than the finite $\mathbf{S}_n$ systems.

On the other hand, a number of applications can be correctly modeled and analyzed in revision sequences of finite length. Uncountable languages, where uncountably many infinite revision sequences necessarily have to be examined, do not play an important and prominent role in most linguistic and philosophical applications.

What does this mean for our question concerning the ontological strength of revision theories? Clearly, one cannot work exclusively in finite systems. Hence, the case in which the models are bound by a countable cardinal number is probably the most important one. For some rare applications it can be necessary to use uncountable models, but they should not occur very often. It is important to emphasize that it is for nearly all reasonable applications possible to work in classical set theory without reference to proper classes. In this respect, revision theories require simply classical set theory from an ontological perspective.

We can conclude that it is possible to restrict the length of the sequences of hypotheses most times. Whether this can be done in a finite way, in a countable or uncountable way, depends on the specific application. In general, that should not create too much trouble. As long as we are not forced to work in revision sequences that are themselves classes and no longer sets, we should accept this. Another problem is whether the picture that is presented by revision theories really captures an intuitive idea of how humans reason when they deal with pathological sentences, belief revisions, or common ground phenomena. Here, the picture of an infinite being that revises hypotheses infinitely often is not very plausible. It seems to be the case that revision theories are relatively good models for a representation of circular phenomena, but they do not tell us anything about the way humans reason about circularity.

In the next subsection, we shall add some remarks concerning higher-order objects that are used in revision theoretic accounts.

10.2.2 Higher-Order Concepts

Theories of grammar are good examples for the development that more sophisticated and better theories can result in an increase of the complexity of these theories. As a side-effect more complex theories are generally using higher-order concepts that did not play any role in the original theories. This becomes obvious by considering the development from classical phrase structure grammars, via Chomsky's Government and Binding Theory to modern theories of grammar like HPSG. Important is to notice that the complexity of these theories increase from contextfree phrase structure grammars via mildly contextsensitive grammars[11] to the non-decidable grammars like HPSG (at least in the non-restricted version of HPSG)[12].

The development from fixed point approaches to revision theories show a very similar behavior. Whereas, fixed point theories as described in Part II of this work are obviously inductively defined, and therefore have complexity Π^1_1, Gupta-Belnap systems are of higher complexity. As we saw, the complexity of validity is Π^1_2. As a consequence of this, we are forced to work with higher-order objects like the collection of all infinite sequences of hypotheses where under certain circumstances every single element of that collection is itself a proper class. Furthermore, the definition of revision theories necessarily requires quantifications over collections of models, higher-order predicates like the property to be a revision sequence or the property of being a reflexive hypothesis in a revision sequence, only to mention some of these higher-order objects. What consequences does the usage of higher-order objects have concerning ontological resources?

First, it should be mentioned that the definitional strength of a theory has important consequences for the theory itself: whereas many first-order theories have the nice property to be decidable this is no longer true for higher-order

[11] Cf. [KoMö99].

[12] Cf. [PoSa94].

theories. Similarly in the revision theoretic case, the infinite Gupta-Belnap systems $\mathbf{S}^*$ and $\mathbf{S}^\#$ are not decidable, and as we examined in Section 8.3, they are even not recursively axiomatizable.

Whether Gupta-Belnap systems are considered as a theory that is appropriate for modeling the truth concept, depends on personal attitudes: either the cognitive truth concept is itself non-decidable or the correspondence between the truth concept and the modeling of it using revision theories is only a weak correspondence, not a strong equivalence. Intuitively, the first reading (namely a weak correspondence between the truth concept and Gupta-Belnap systems) is a more attractive reading, because we need not to explain why cognitive abilities exceed the principal border of computability. In the second reading, we need to give reasons how this can be possible at all. Whether a theory exists that yields results as good as revision theories, but uses weaker resources (in the best case decidable resources) remains an open question.

Another question concerning higher-order objects (as for example transfinite revision sequences in comparison with finite revision sequences) is how the difference between finitely many revisions and a transfinite number of revisions can create a difference at all because human capacities and human reasoning are limited to a finite number of revisions. In other words: it is absurd to consider the possibility that human reasoning concerning the pathological behavior of a sentence (or a set of sentences) is performed by an infinite number of revisions. Why is there a difference between the systems $\mathbf{S}_n$ and $\mathbf{S}^*$ at all (and similarly between the $\mathbf{S}_n$ systems and the system $\mathbf{S}^\#$ as well)? And why can we easily understand that difference?

A possible explanation is analogous to an explanation why human cognitive capacities are able to deal with infinite objects in mathematics: we can transcend the reasoning about infinite objects by specifying properties, shortcuts, and variables for infinite processes. In fact, human capacity is able to reason with finite means about infinite objects and infinite operations. As well as we can transcend infinity, we can transcend the concept of truth (which is not a finite concept) with finite means. Clearly, this is far away from a concise thesis how human capacities can deal with infinity, but it is an intuitively plausible claim that there are similarities between these two capacities.

10.2.3 Sentences versus Propositions

As in Subsection 6.4.3, we can ask which entity is the bearer of truth. Standardly, in revision theories as well as in fixed point approaches, it is assumed that the bearers of truth are sentences. This makes it easier to apply Gupta-Belnap systems to a particular phenomenon provided contexts need not to be considered. Whether this strategy is appropriate from a wider and more general perspective remains an open question.

As we saw in the considerations concerning fixed point approaches and as we will see later in Chapter 16, there are strong reasons for assuming that an adequacy condition for a formal theory of circularity is the possibility to extend that theory to an account that can represent propositions as well. Whereas this is a problem for fixed point theories, we saw in Section 9.4 that in principal there are no a priori reasons for the impossibility of such an extension in the revision theoretic case. Clearly, the problem of how to define such a theory in detail, as well as the fact that there are a lot of open problems concerning a general theory of propositions, does not simplify the situation. Whereas the fixed point approach seems to prevent an introduction of propositions, Gupta-Belnap systems do not block this extension on a principal level. In this respect, revision theories are the more flexible accounts towards a theory of circularity.

10.3 History

The claim that the strengthened Liar sentence provides a problem for revision theories was claimed in [Ca86], [Pr87], [Bre92], and [Ma96]. In Gupta and Belnap's monography [GuBe93], one can find a good discussion of this point. Our presentation of the strengthened Liar here is very close to the discussion there. Problems concerning the Truth-teller circle and the modified Gupta puzzle were mentioned in [Ya93]. A proposal for a solution was described in [Ch96]. In [GuBe93], a discussion concerning logical tautologies and the evaluation of these tautologies in revision theories can be found. Which ontological entities must be assumed in order to develop a particular theory as well as the question which entity is the bearer of truth are prominent questions in the field of truth theories in general. With respect to revision theories, the author does not know any discussion of these aspects, although they are important because of the non-standard objects used in revision theoretic approaches.

Part IV

An Alternative to ZFC and the Theory of Coalgebras

Chapter 11

The Non-well-founded Universe

Classical ZFC set theory was developed in order to avoid certain objects that yield inconsistencies. The classical example is the restriction of the full comprehension axiom. Another example is the foundation axiom. This axiom prohibits the existence of sets that contain themselves. For example, a set a that satisfies the condition $a = \{a\}$ simply does not exist in ZFC. It is not a priori clear why this restriction is appropriate for a reliable and expressively powerful theory of sets. Clearly, there is no simple answer or argument for or against the attempt to develop a theory of non-well-founded sets.[1] We will be able to give some arguments concerning this more philosophical question later. In this chapter, we will develop a set theory that allows the existence of sets that contain themselves as elements. Consistency of the resulting theory will be preserved relative to the consistency of ZFC. In other words: if ZFC is consistent, then the new non-well-founded set theory is consistent, too.

We will be quite explicit in formulating the relevant notions and concepts in the following. One reason is to provide the necessary prerequisites for the further development in this part of the present work. Another reason is the fact that most readers are probably not very familiar with non-well-founded set theory. Nevertheless, a full discussion of all relevant implications of the development concerning the basic theory of non-well-founded set theory is not possible here.

11.1 Preliminaries

In this section, we will consider some elementary definitions and facts concerning ordinary set theory with urelements. Urelements - considered as objects that are not sets themselves but can be used in order to generate sets - are rather uncommon in classical set theory, because they are nearly almost superfluous to provide a basis for classical mathematics. The objects mathematicians commonly use in their theories are usually entities like numbers, spaces, fields,

[1] We sometimes call non-well-founded sets hypersets. The two notions are assumed to be coextensional.

functions etc. For all these objects it is not necessary to introduce urelements. In general, these objects can be represented using sets that are constructed from the empty set. For our purposes this is not sufficient, because we want to model 'real-world' phenomena. That's the reason why we will work in set theory with urelements.

11.1.1 Basics

We begin our examination with some remarks concerning ZFC set theory. A full discussion of the axioms of ZFC (Zermelo-Fraenkel set theory with the axiom of choice) is not possible in this chapter. Only the most important properties and constructions will be mentioned. If the assiduous reader is interested in further developments and properties of classical ZFC set theory, he is encouraged to consult well-known introductions into and monographs of the field of axiomatic set theory. Examples of readable books about ZFC are [Su60, De93], and [Mo94]. The standard (but quite advanced) reference for ZFC is [Je78]. Our representation of classical set theory as well as our treatment of the theory of hypersets is very close to the monograph [BarMo96]. This work presents the important ideas of non-well-founded set theory and additionally contains a lot of applications in the fields of mathematics, computer science, philosophy, and logic.

We need to introduce some basic definitions. An important concept of set theory is the idea of operations on sets. It is not sufficient simply to speak about sets. At least as important as to know which kind of sets there are is to know which kind of operations are defined on these sets. Operations enable us to generate new sets, if certain other sets are given. One of the basic concepts is the concept of an ordered pair of two sets. The following definition makes this concept precise.

Definition 11.1.1 *(i) An ordered pair $\langle a, b\rangle$ where a and b are not necessarily sets is defined as the following set theoretic object:*[2]

$$\langle a, b\rangle = \{\{a\}, \{a, b\}\}$$

(ii) We define equality of two ordered pairs $\langle a, b\rangle$ and $\langle c, d\rangle$, provided that a, b, c, and d are sets as follows:

$$\langle a, b\rangle = \langle c, d\rangle \;\Leftrightarrow\; a = c \textit{ and } b = d$$

We can interpret the concept of an ordered pair as a possibility to code complex expressions set theoretically. Notice that the set coding an ordered pair has a deep structure and is not flat.

It is possible to define more complex mathematical objects, such as relations and functions using ordered pairs. We can view functions and relations as sets consisting of pairs that satisfy certain conditions in the case of functions.

[2] An equivalent representation is: $\langle a, b\rangle = \{\{b\}, \{a, b\}\}$.

Definition 11.1.2 *(i) A relation R is a set X, such that every element of X is an ordered pair of sets.*
(ii) A function f is a set Y, such that every element of Y is an ordered pair of sets with the additional property: if $\langle a, b\rangle \in Y$ and $\langle a, c\rangle \in Y$, then it holds: $b = c$.

In a more familiar notation Definition 11.1.2 means: if $f(a) = b$ and simultaneously $f(a) = c$ then $b = c$, i.e. every argument in the domain of a function f is mapped via f to a unique value in the range of f. That is the standard definition of a function.

The presented definitions of functions and relations are relatively abstract, because there is no domain or range explicitly specified. Clearly, the domain can be reconstructed by collecting the first argument of the pairs. Similarly this can be done for the range. In classical analysis, functions are usually defined on numbers. The basis on which arbitrary numbers can be coded as set theoretical objects are natural numbers. John von Neumann defined natural numbers as transitive, well-founded, and linear sets. His construction can be viewed (for the finite case) by the following 'definition':

$$\begin{array}{lllll}
0 & = & \emptyset & & \\
1 & = & \{0\} & = & \{\emptyset\} \\
2 & = & \{0,1\} & = & \{\emptyset,\{\emptyset\}\} \\
\vdots & \vdots & \vdots & \vdots & \vdots
\end{array}$$

The general idea of the construction is the following one: take a particular set x (in our case the empty set) and associate it with 0. Then, the successor number of x is calculated by the application of the power set operation. That process can be repeated. An alternative definition of the natural numbers is the following construction (that is due to Zermelo):

$$\begin{array}{llll}
0 = \emptyset & & & \\
1 = \{0\} = \{\emptyset\} & & & \\
2 = \{\{0\}\} = \{\{\emptyset\}\} & & & \\
\vdots & \vdots & \vdots & \vdots
\end{array}$$

We do not consider the transfinite case of von Neumann's (or Zermelo's) construction. Natural numbers can be extended to the theory of ordinal numbers. Similarly to the case of natural numbers it is possible to define an arithmetic on ordinals. Good introductions into the subject of ordinals are the books [Si58, Je78], and [Su60].

Given the natural numbers, how can we define integers as set theoretic objects? The standard account is to associate integers with pairs of natural numbers. The following definition makes this precise.

Definition 11.1.3 *(i) Assume the natural numbers $\mathbb{N}$ are given by von Neumann's construction. Consider the collection of all pairs $\mathbb{Z} = \{\langle a,b\rangle \mid a \in \mathbb{N} \ \wedge \ b \in \mathbb{N}\}$. Define an equivalence relation $\equiv$ on pairs $\langle a,b\rangle \in \mathbb{Z}$ as follows:*

$$\langle a,b\rangle \equiv \langle c,d\rangle \quad \iff \quad \begin{cases} a-b=c-d & : \quad a \geq b \wedge c \geq d \\ b-a=d-c & : \quad a < b \wedge c < d \end{cases}$$

(ii) Define an operation $+$ on $\mathbb{Z}$ according to the following condition:

$$\langle a,b\rangle + \langle c,d\rangle = \langle a+c, b+d\rangle$$

Remark 11.1.1 (i) One can easily check that the defined relation $\equiv$ is an equivalence relation on $\mathbb{Z}$. Furthermore, addition $+$ is well-defined and has precisely the intended meaning.

(ii) The above construction can be extended to rational numbers as well, where pairs of integers represent rational numbers. This is a standard and well-known construction and will not be introduced here.

The operations Cartesian product, union, intersection, power set operation, and set theoretical difference are defined in Chapter 1. Some of these operations will be reconsidered in Subsection 11.1.3 when we will examine the axioms of set theory. We refer the reader to this subsection for further information. We mention only the general idea of such a justification. The power set operation and the union operation can be justified directly by the corresponding axioms of set theory. For other operations we need a combination of several axioms. For example, the Cartesian product is justified because of the power set axiom, the paring axiom, and the separation axiom. More precisely, the following relation holds:

$$A \times B \subseteq \wp(\wp(A \cup B))$$

Because of its importance we mention a special union operation where we keep in mind from which set one got a particular element: the so-called disjoint union of two (or more) sets. The motivation for this kind of operation is that it is sometimes necessary to be able to reconstruct whether an element of the union of two sets a and b originates from a or from b. We can model this idea of a disjoint union of two sets a and b by the following construction:[3]

$$a \oplus b = (\{0\} \times a) \cup (\{1\} \times b)$$

In the next subsection, we will introduce transitive sets. This concept will occur in many applications of set theory.

[3] Notice that the Cartesian product of sets is not associative.

11.1.2 Transitivity

A very important concept in set theory (as well as in other branches of mathematics) is transitivity. In our context, transitive sets will be considered. Additionally, we need the concept of the transitive closure of a set. The following definition introduces the concept of a transitive set.

Definition 11.1.4 *Assume a set a is given. We call a transitive, if the following condition holds:*

$$(\forall b)(\forall c) : (c \in b \ \wedge \ b \in a \ \rightarrow \ c \in a)$$

There are equivalent definitions specifying this property of sets. To become familiar with the concept of transitivity we state the following proposition that associates three variants of this property.

Proposition 11.1.5 *Assume a is a set. Then the following properties (i) - (iii) are equivalent:*

(i) $(\forall b)(\forall c)(c \in b \ \wedge \ b \in a \ \rightarrow \ c \in a)$
(ii) $(\forall b)(b \in a \ \rightarrow \ b \in \wp(a))$
(iii) $\bigcup a \subseteq a$

Proof: "(i) ⇔ (ii)": The following equivalences show that (i) and (ii) are equivalent.

$$\begin{array}{rl} & (\forall b)(\forall c)(c \in b \wedge b \in a \rightarrow c \in a) \\ \Leftrightarrow & (\forall b)(\forall c)[(b \in a) \rightarrow ((c \in b) \rightarrow (c \in a))] \\ \Leftrightarrow & (\forall b)[(b \in a) \rightarrow (\forall c)((c \in b) \rightarrow (c \in a))] \\ \Leftrightarrow & (\forall b)[(b \in a) \rightarrow (b \subseteq a)] \\ \Leftrightarrow & (\forall b)[(b \in a) \rightarrow (b \in \wp(a))] \end{array}$$

"(i) ⇔ (iii)" The following equivalences show that (i) and (iii) are equivalent.

$$\begin{array}{rl} & (\forall b)(\forall c)(c \in b \ \wedge \ b \in a \ \rightarrow \ c \in a) \\ \Leftrightarrow & (\forall b)[(b \in a) \rightarrow (\forall c)((c \in b) \rightarrow (c \in a))] \\ \Leftrightarrow & (\forall b)[(b \in a) \rightarrow \{c \mid c \in b\} \subseteq a] \\ \Leftrightarrow & \{c \mid c \in b \text{ for } b \in a\} \subseteq a \\ \Leftrightarrow & \bigcup a \subseteq a \end{array}$$

This suffices to show the claim of the proposition. q.e.d.

We will introduce an important operation on sets which is closely related to the transitivity of sets, namely the transitive closure of a given set a.[4] The

[4] Instead of transitive closure, sometimes authors use the notion 'transitive hull'.

transitive closure of a set a (usually denoted by $TC(a)$) is the smallest transitive set which includes a. Intuitively, there are two possibilities of defining the transitive closure. The first possibility is to define it 'from above', the second possibility is to define it 'from bottom up'. The following definition corresponds to the idea to define the transitive closure from bottom up using essentially an inductive process.

Definition 11.1.6 *The transitive closure $TC(a)$ of a set a is defined as follows:*

$$TC(a) = \bigcup\{a, \bigcup a, \bigcup\bigcup a, \dots\}$$

In order to associate the definition of transitive closure with Definition 11.1.6, we state a further proposition.

Proposition 11.1.7 *Assume a is a set. Then it holds: a is transitive iff a satisfies the following condition:*

$$TC(a) = \bigcap\{x \mid a \subseteq x \ \wedge \ x \textit{ is transitive}\}$$

Proof: Let a be an arbitrary set. First, we show that

$$TC(a) = \bigcap\{x \mid a \subseteq x \ \wedge \ x \text{ is transitive}\}$$

implies that a is transitive. An easy consideration shows that the intersection of transitive sets preserves transitivity. Therefore, if there are any transitive sets which include a, then the above condition guarantees that there is a smallest transitive set including a. Therefore, we have to show that there is at least one such transitive set which includes a. To show this, assume $x \in \bigcup\{a, \bigcup a, \bigcup\bigcup a, \dots\}$. Then, by definition there is a natural number $n \in \omega$, such that $x \in \bigcup^n a$. Let y be in x. Then, it holds $y \in \bigcup^{n+1} a$. Hence, it holds $y \in \bigcup\{a, \bigcup a, \bigcup\bigcup a, \dots\}$ as well. In other words: $TC(a) = \bigcup\{a, \bigcup a, \bigcup\bigcup a, \dots\}$ is transitive. Because $a \subseteq \bigcup\{a, \bigcup a, \bigcup\bigcup a, \dots\}$, it must be the case that $TC(a)$ is a transitive set which includes a. This shows that the above construction is well-defined.
In order to show the other direction, it remains to show that $\bigcup\{a, \bigcup a, \bigcup\bigcup a, ...\}$ is the smallest transitive set including a. This can be proven by an easy induction: assume b is a transitive set which includes a. For $n = 0$ we have: $a \subseteq b$ by assumption. Assume it holds for $n \in \omega$: $\bigcup^n a \subseteq b$, then $\bigcup(\bigcup^n a) \subseteq \bigcup b \subseteq b$. Therefore, for all $n \in \omega$ we have $\bigcup^n a \subseteq b$. Conclude: $TC(a) = \bigcup\{a, \bigcup a, \bigcup\bigcup a, \dots\} \subseteq b$ q.e.d.

The non-well-founded set theory we will describe in this chapter has an additional feature that is in general not included in ordinary set theory: we are enabled to form sets that consist of objects that are non-sets themselves. All objects that can be members of a set but are themselves non-sets, are called urelements. We denote the proper class of these objects by $\mathcal{U}$. The question arises whether it is necessary to assume the existence of a proper class of urelements in order to develop the intended theory of non-well-founded sets. In fact, there is no a priori reason for this assumption. It is simply easier to have in all possible situations enough urelements left that can be used for further constructions.[5]

Because of the existence of urelements we have that not every element of a given set is necessarily itself a set. This is quite unfamiliar at first sight, but it does not cause any structural problems. Furthermore, it is useful to define a function (operation) which provides new urelements. Given a set a and a collection of urelements $b \in \mathcal{U}$ the expression $\texttt{new}(a, b)$ denotes a new urelement that is not contained in b. Additionally, we will require that this function `new` is injective with respect to the first argument. Formally, we can state this precisely in the following definition.

Definition 11.1.8 *There is a function new(a,b), such that the following two conditions (i) and (ii) hold:*

(i) for all sets a and b with $b \subseteq \mathcal{U}$: $new(a, b) \in \mathcal{U} - b$
(ii) For all sets a, a', b with $a \neq a'$ and $b \subseteq \mathcal{U}$: $new(a, b) \neq new(a', b)$

Urelements are not standard in classical ZFC set theory. Justifications for the introduction of urelements will be given by applications of non-well-founded set theory on the one hand and by the development and generalization of the theory of hypersets in this chapter and in later chapters on the other. From an intuitive point of view it is clear that we would like to be able to model 'real-world' phenomena and not only mathematical entities. Because of this fact urelements are needed.

Given a set a it is sometimes important to know, which urelements are 'involved' in a. Therefore, we introduce the concept 'support of a given set a', which is a first application of the transitive closure of a set and can best be described as follows: collect all elements that are somehow embedded in a. This collection is called support of a. The support operation on sets will be used quite extensively in the following.

Definition 11.1.9 *The support of a set a, is defined as follows:*

$$support(a) = TC(a) \ \cap \ \mathcal{U}$$

Definition 11.1.9 allows us to specify the urelements that are embedded in a particular set. Notice that in set theory without urelements it is not necessary to introduce a corresponding operation, because all objects embedded in a given set are necessarily sets.

The next subsection gives an overview of the axioms of ZFC. Because we develop a universe that includes non-well-founded sets our axiomatic theory will differ slightly from ZFC set theory.

[5] A further reason is to be able to speak about the real world in possible applications of the theory of hyperets.

11.1.3 Axioms

The main interest in this chapter is the development of a theory of non-well-founded sets. Therefore, the foundation axiom (that holds in ZFC) must be replaced, because it blocks the existence of reflexive sets, i.e. sets that are included in themselves. The strategy we will follow is to cancel this axiom and to introduce a certain kind of anti-foundation axiom. This was originally the account in [Ac88] (although Aczel proposed a variety of different versions of anti-foundation axioms). Before we will develop this idea further some facts concerning the classical foundation axiom and classical ZFC set theory must be examined.

We begin with the foundation axiom and some properties of this axiom. The idea is to require for every set in the universe that it contains a minimal element with respect to the $\in$ relation. Usually, the foundation axiom is formulated as follows:

$$\forall a(a \neq \emptyset \;\rightarrow\; \exists x(x \in a \;\wedge\; \forall y \in a(y \in x \;\rightarrow\; y \notin a)))$$

An equivalent representation of the foundation axiom can be given by the following formula:[6]

$$\forall a(a \neq \emptyset \;\rightarrow\; \exists x(x \in a \;\wedge\; x \cap a \;=\; \emptyset))$$

Intuitively, a set a is well-founded if it contains an element x that is in a certain sense minimal. We cannot find another element y of x that is contained in a. An immediate consequence of this fact is that there is no descending sequence of elements of a set a which is infinite. Assuming the axiom of foundation the following formula holds:

$$\neg(\exists a \exists a_1 \exists a_2 \ldots : (\ldots \in a_n \in a_{n-1} \in \ldots \in a_2 \in a_1 \in a))$$

Notice that every circle of the form $a \in b_1 \in b_2 \in \ldots \in b_n \in a$ implies that there is a reflexive set. Using the transitivity of $\in$ we can deduce immediately that it holds $a \in a$ which is impossible provided the foundation axiom holds. Hence, under the assumption that foundation is true we can prove the following claim:

$$\neg(\exists a \exists a_1 \exists a_2 \ldots \exists a_n : (\ldots \in a_n \in a_{n-1} \in \ldots \in a_1 \in a \in a_n \in \ldots))$$

There are a few more facts that can be proven using the foundation axiom. The following properties give an idea which statements are directly implied by the foundation axiom. Notice that in the non-well-founded set theory we will develop later these formulas do not hold in general.

Fact 11.1.10 *Provided the foundation axiom holds the following statements are provable.*

[6] The equivalence of the two formulations can be easily checked.

(i) $\forall a(a \notin a)$
(ii) $\neg\exists a\exists b(a \in b \wedge b \in a)$
(iii) $\forall a(a \subseteq a \times a \rightarrow a = \emptyset)$
(iv) $\forall a\forall b\forall c(a = \langle b, c\rangle \rightarrow a \neq b \wedge a \neq c \wedge a \notin b \wedge a \notin c)$

Proof: Proofs of the statements (i) - (iv) can be found in most textbooks of set theory. We do not show these facts here. q.e.d.

In the following, we will consider other axioms of set theory. The basic motivation of the axiomatic approach of ZFC set theory is the attempt to give an answer to the question: 'What are sets?' Far from claiming to give a satisfactory answer to this deep and controversial question, nevertheless, the axiomatic approach fits into a rational discussion of this question. In the following, we will give an overview which axioms are used in set theory and what the motivation for such axioms really is.

(i) **Axiom of Extensionality:** Assume two sets a and b are given. We would like to specify criteria in order to be able to decide, whether the two sets are equal, or not. Intuitively, two sets are equal, if they contain the same elements. That is exactly what is stated in the so-called axiom of extensionality.

$$(\forall a)(\forall b)(\forall c)(a = b \leftrightarrow (c \in a \leftrightarrow c \in b))$$

It should be mentioned that the equivalence relation in the above axiom is not necessary, because the left-to-right implication is an axiom of ordinary first order logic. Therefore, in many textbooks this axiom is stated in conditional form. We use the biconditional form in order to make clear the intuition behind this axiom.

(ii) **Pairing Axiom:** Given two sets a and b the pairing axiom is a principle which claims the existence of a set that has exactly sets a and b as elements. The formal axiom can be represented as follows:

$$(\forall a)(\forall b)(\exists c)(\forall d)(d \in c \leftrightarrow d = a \vee d = b)$$

(iii) **Union Axiom:** In order to ensure that the naive union of two sets a and b is justified, we state a more general principle, namely the axiom of union for arbitrary sets. Formally, it can be represented as follows:

$$(\forall a)(\exists b)(\forall c)(\forall d)(d \in c \wedge c \in a \leftrightarrow d \in b)$$

An obvious consequence of the general form of the union axiom is the appropriateness to model naive union of two sets $a \cup b$. This can easily by checked by the following considerations:

$$a \cup b \;=\; \{x \mid x \in a \vee x \in b\} \;=\; \{x \mid x \in y \text{ for } y = a \vee y = b\} \;=$$
$$=\; \{x \mid x \in y \text{ for } y \in \{a, b\}\} \;=\; \bigcup\{a, b\}$$

(iv) **Power set Axiom:** To justify the existence of the power set of a given set a we state the power set axiom as follows:

$$(\forall a)(\exists b)(\forall c)(c \subseteq a \;\leftrightarrow\; c \in b)$$

The power set axiom is a very strong set theoretic principle. It is crucially used to define the universe of the (so-called) constructible sets. We also saw this principle in the construction of the natural numbers above. In terms of the notion of cardinality, iterated applications of the power set axiom strictly increases the cardinality of a set. That is a consequence of Cantor's theorem.

(v) **Axiom of Separation:** The separation axiom enables us to define naive intersection, and the subtraction of two sets. Additionally, we can state the existence of the empty set, assuming that there is already a set in the universe. (The idea that there is a set at all is captured by the axiom of the existence of an infinite set). For a given formula $\phi(b, p)$ the separation axiom can be formulated as follows:

$$(\forall A)(\forall p)(\exists B)(\forall b)(b \in B \;\leftrightarrow\; b \in A \;\wedge\; \phi(b, p))$$

The separation axiom can be considered as a weak form of comprehension. A set b is an element of a set B, if a condition $\phi(b, p)$ is satisfied and additionally b is taken from another set. Hence, it cannot happen that the collection of all sets satisfying a condition $\phi(b, p)$ becomes class-sized.

(vi) **Axiom of Replacement:** The axiom of replacement has the following form:

$$(\forall a)(\forall b)(\forall c) :\; (\phi(a, b, p) \;\wedge\; \phi(a, c, p) \;\rightarrow\; b = c) \;\rightarrow$$
$$(\forall A)(\exists B)(\forall b) :\; (b \in B \;\leftrightarrow\; (\exists a \in A) : \phi(a, b, p))$$

The axiom of replacement can be interpreted as follows: if a function ϕ is given and we restrict the domain of ϕ to a certain set, then the range is also a set.

(vii) **Existence of an infinite set:** Still, we do not know whether there exists a set at all. The axioms we considered so far have the form: if one (or more) set(s) exist(s), then another set can be constructed. But we do not know, whether there exists a set at all. The next axiom - the axiom of the existence of an infinite set - gives us an affirmative answer to this problem.

$$(\exists a)(a \neq \emptyset \wedge \forall b(b \in a \rightarrow b \cup \{b\} \in a))$$

In this axiom, crucially a notion of finiteness is used which is borrowed by von Neumann's definition of the natural numbers. In order to specify an infinite set, one has to use one or the other notion of finiteness. Notice that the axiom of the existence of an infinite set does not uniquely define a particular set.

(viii) **Axiom of Foundation:** For a discussion of the foundation axiom (sometimes this axiom is also called axiom of regularity), the reader is referred to the remarks above. For completeness we state this axiom again:

$$\forall a(a \neq \emptyset \ \rightarrow \ \exists x(x \in a \ \wedge \ \forall y \in a(y \in x \rightarrow y \notin a)))$$

(ix) **Axiom of Choice:** In ZFC, the axiom of choice ensures that for every set a there exists a function f, such that for every non empty subset b of a it holds: $f(b) \in b$ (f is called a choice function). Formally we can state this as follows:

$$(\forall a)(\exists f)(\forall b)(b \subseteq a \ \rightarrow \ f(b) \in b)$$

The axiom of choice was a strongly discussed axiom in the history of mathematics. Reasons for this controversial discussion are some counterintuitive consequences in certain applications of this axiom. Additionally, the axiom of choice is a highly non-constructive principle. We do not go into details concerning the properties and consequences of the axiom of choice. We mention three principles (b) - (d) that are equivalent to this axiom.

The following properties are equivalent:

(a) $(\forall a)(\exists f)(\forall b)(b \subseteq a \rightarrow f(b) \in b)$

(b) Every set can be well-ordered.

(c) For all sets a, b, c it holds: $|a| < |b| \ \vee \ |a| = |b| \ \vee \ |b| < |a|$.

(d) If $a \neq 0$ and if the supremum of each non-empty chain which is a subset of a is in a, then a has a maximal element.

A nice and quite explicit overview of the history of the axiom of choice can be found in [Mo82].

(x) **Urelements:** The axiom of the existence of urelements is not a standard axiom in ordinary ZFC set theory. Using the function `new` we require the existence of a proper class of urelements. Urelements are objects that are neither sets nor classes. In particular, urelements do not contain any other elements. Urelements do not have any inner structure, they are atomic.

> The axiom below states the property that an urelement does not contain any other object. (The expression $\mathcal{U}(p)$ denotes the statement that p is a urelement.)
>
> $$(\forall p)(\forall q)(\mathcal{U}(p) \;\rightarrow\; q \notin p)$$

These remarks finishes the overview of the axioms of ZFC. Much more could be said concerning the properties and consequences of these axioms. Furthermore a discussion could be added concerning the reliability of ZFC. The interested reader is referred to the standard literature about this topic.

11.1.4 Classes and Sets

ZFC set theory provides a good foundation for mathematical theories. It is the best 'ontological' basis for formal sciences that is known and sufficiently developed. Additionally, the success of the foundations of mathematics using axiomatic set theory is a strong argument for the reliability of this approach. Unfortunately, there is a price to pay for the elimination of the paradoxes in set theory. The universe is limited to certain 'small collections' of objects. In other words, it is not allowed that the collections become too large. These larger collections, namely proper classes, are excluded in the set theoretical universe. We will add some remarks concerning classes here.

A relatively good approximation for the well-known and important distinction between sets and classes can be achieved by considering large sets as classes. For example, the famous Russell set, given by $X = \{x \mid x \notin x\}$ which collects all sets that are not reflexive[7] cannot be a set itself, because if it were, then immediately one would run into contradictions. We already presented Russell's argument in Chapter 2. Historically, this was a very important insight, because it became clear that naive set theory with full comprehension is not a reliable foundation of mathematical theories. The consequence was that Zermelo and Fraenkel developed the axiomatic set theory ZFC.

The standard solution of this inconsistency of naive set theory with full comprehension is the distinction between sets and classes. Examples of proper classes can easily be given: the collection of all non-reflexive sets, the collection of all ordinal numbers, or the collection of all singletons.[8] Other examples are the collection of all Banach spaces, or the collection of all fields etc. It is important to recognize that this distinction between sets and classes is not a result of work about set theory naively, but a consequence of a concept 'set' that is restricted and clearly specified by the axioms of ZFC set theory.

The following remarks summarizes some consequences of the axiomatic approach of set theory, which are closely related to the class versus set distinction.

[7] A set a is called reflexive, if a is a member of itself or, in other words, if a satisfies the condition: $a \in a$.

[8] A singelton is a set which contains only one element.

Remark 11.1.2 (i) Every predicate determines a class but not every predicate determines a set. The idea of ZFC is it to restrict the 'ordinary' axiom of comprehension to an axiom where every element of the new set determined by a condition ϕ must be taken from a given set a. Hence, every set b specified by the separation axiom is a subset of a. Therefore, sets defined by comprehension cannot become too large in ZFC.

(ii) We already have seen some closure properties of ZFC. For example, sets are closed under taking images of functions. There are some other closure operations: taking the union of a set, the intersection, or the difference of two sets (these are direct consequences of the axiom of union, or the separation schema, respectively), or building the pair of two sets (pairing axiom). Even if we apply the power set operation to a set a we get a new set $\wp(a)$ as output. We refer the reader to textbooks of set theory for a more elaborated discussion of this point.

(iii) A consequence of the union axiom is it that every member of a set is a set. For example: if $\{a\}$ is set, then a is a set, too.[9] This is in a certain sense counterintuitive, if one considers singletons, i.e. collections that contain only one element. For example, $\{\mathcal{V}\}$, the one-element collection of the collection of all sets is intuitively a small collection, simply because it contains only one element. But $\{\mathcal{V}\}$ cannot be a set, because of the union axiom.

(iv) In general, classes do not play a very prominent role in mathematics. Most mathematical objects are sets of the set theoretic universe. Important exceptions are certain collections used in category theory. Here, one distinguishes between small categories (set-sized categories) and large categories (class-sized categories). Although category theory plays more and more an important role in various applications most times large categories are not needed. We will consider some aspects of category theory in Chapter 12 of this work.

We finish this section with the remark that the presented axioms of ZFC should not be considered as the one and only possibility for axiomatizing set theory. We will see in the following sections, how to generate a quite different set theoretic universe. This new universe contains the sets of ZFC but extents the collection of these sets by adding reflexive sets (hypersets). We can speak of an enlargement of classical ZFC set theory.

11.2 The Anti-Foundation Axiom

There are different ways to introduce hypersets. Peter Aczel uses in [Ac88] labeled (decorated) graphs for stating the anti-foundation axiom. Modulo some equivalent representations of this axiom one can formulate it - in the spirit of Aczel's account - as follows: every graph can be uniquely decorated and every

[9] Assume for the moment that urelements are not available.

decorated graph corresponds to a set in the non-well-founded universe. Despite the fact that this introduction of the anti-foundation axiom is an intuitive graphical representation, we will not adopt it here. The reason for choosing a different idea is the fact that the extension of the set theoretical universe by solutions of set theoretical equations is for most people easier to grasp. Clearly, this may change from person to person.

We choose the account explicitly developed in [BarMo96]. A natural and well-known construction often used in mathematics to introduce new objects can also help to introduce non-well-founded sets. This construction corresponds to the idea that certain set theoretical equations have a unique solution. We prefer this approach to introduce hypersets, because of the corresponding constructions in elementary number theory and the obvious simplicity of this account.

First, we need to specify the notion of a set theoretical systems of equations. According to [BarMo96], we can distinguish four different types of such systems. The next definition introduces these four types. Later we will see that the axiom that one type of system has a unique solution is equivalent to the axiom that another type of system has a unique solution.

Definition 11.2.1 *(i) A flat (set theoretic) system of equations is a triple $\mathcal{E} = \langle X, A, e\rangle$ where $X \subseteq \mathcal{U}$ is a set of urelements,*[10] *A is a set of atoms,*[11] *X and A are distinct (i.e. $A \cap X = \emptyset$), and e is function, specified as follows: $e : X \longrightarrow \wp(X \cup A)$.*

(ii) A generalized flat systems of equations is a triple $\mathcal{E} = \langle X, A, e\rangle$ where X and A are arbitrary sets with the additional property that $X \cap A = \emptyset$ and $e : X \longrightarrow \wp(X \cup A)$.

(iii) Let a be an arbitrary set. A canonical system of equations for a is a triple $\mathcal{E} = \langle TC(\{a\}) - support(a), support(a), e\rangle$ where $e(x) = x$ for all $x \in TC(\{a\}) - support(a)$, i.e.: e is the identity function on the set $TC(\{a\}) - support(a)$.

(iv) A general system of equations is a triple $\mathcal{E} = \langle X, A, e\rangle$ where $X \subseteq \mathcal{U}$, $A \subseteq \mathcal{U}$, $X \cap A = \emptyset$, and e is a function mapping X into $\mathcal{V}_{afa}[X \cup A]$.

Notice that in Definition 11.2.1(iv) we used the concept $\mathcal{V}_{afa}[X \cup A]$. We do not know yet what the non-well-founded universe relative to the urelements $X \cup A$ really is. We will get more information soon. In order to give a complete list of the different forms of systems of equations, we included the general system above. We need to define the concept of a solution of these systems.

[10] X should be understood as the set of indeterminates of the system.

[11] Atoms are non-structured objects relative to our system. Although we will consider in most cases the set of atoms A as a subset of $\mathcal{U}$, we permit (in flat systems and generalized systems) that A is an arbitrary set.

The following definition provides additional information concerning solutions of systems of equations.

Definition 11.2.2 *(i) Assume $\mathcal{E} = \langle X, A, e\rangle$ is a flat system of equations. A solution to $\mathcal{E}$ is a function s with dom$(s) = X$, such that for each $x \in X$ it holds:*

$$s(x) = \{s(y) \mid y \in e(x) \cap X\} \cup (e(x) \cap A)$$

(ii) Assume $\mathcal{E} = \langle X, A, e\rangle$ is a generalized flat system. A solution to $\mathcal{E}$ is a function s with dom$(s) = X$, such that for each $x \in X$, it holds:

$$s(x) = \{s(y) \mid y \in e(x) \cap X\} \cup (e(x) \cap A)$$

(iii) Let $\mathcal{E} = \langle TC(\{a\}) - support(a), support(a), e\rangle$ be a canonical system of equations. A solution to $\mathcal{E}$ is the identity function on $TC(\{a\}) - support(a)$.

We add some remarks concerning the above definition.

Remark 11.2.1 (i) A solution s of a general system of equations $\mathcal{E}$ will be introduced later. The reason for this is the fact that we need additional concepts like the concept of corecursive substitution in order to define a reasonable notion of a solution.

(ii) Notice that the definitions of solutions for flat systems and generalized flat systems are completely equal. Consequently, it is not necessary to interpret the indeterminates $x \in X$ of a system as urelements. Ordinary sets can do the job equally well.

(iii) Notice further that all described systems cannot solve equations of the form $a = \wp(a)$ where a is a set. According to Cantor's theorem there is no set a that satisfies this equation. We will see that certain theorems, like Cantor's theorem, do hold in non-well-founded set theory, too.

A crucial concept which is important for many theorems in the further elaboration of the theory is the concept of a bisimulation. We can distinguish several forms of bisimulations: for example, we can consider bisimulations between systems of equations or bisimulations on sets. The next definition introduces the concept of a bisimulation between systems of equations. The definition of bisimulations on sets will be stated later. Other forms of bisimulations will be introduced in Chapter 13.

Definition 11.2.3 *(i) Let $\mathcal{E} = \langle X, A, e\rangle$ and $\mathcal{E}' = \langle X', A', e'\rangle$ be two flat systems of equations. A bisimulation relation R between $\mathcal{E}$ and $\mathcal{E}'$ is a relation*

$R \subseteq X \times X'$, *such that the following three conditions (a)-(c) hold:*

(a) $xRx' \ \rightarrow \ (\forall y \in e(x) \cap X)(\exists y' \in e'(x') \cap X') : yRy'$
(b) $xRx' \ \rightarrow \ (\forall y' \in e'(x') \cap X')(\exists y \in e(x) \cap X) : yRy'$
(c) $xRx' \ \rightarrow \ e(x) \cap A = e'(x') \cap A$

(ii) Let $\mathcal{E} = \langle X, A, e\rangle$ *and* $\mathcal{E}' = \langle X', A', e'\rangle$ *be two flat systems of equations.* $\mathcal{E}$ *and* $\mathcal{E}'$ *bisimulate, if there is a bisimulation relation* R *between* $\mathcal{E}$ *and* $\mathcal{E}'$ *and additionally conditions (a) and (b) hold:*

(a) $(\forall x \in X)(\exists x' \in X') : xRx'$
(b) $(\forall x' \in X')(\exists x \in X) : xRx'$

The idea of the concept of bisimulations is that mathematical objects that are structurally similar can be associated via a bisimulation relation R. Notice that a bisimulation between two systems is in general much weaker than an isomorphism between these systems. For example, a bisimulation relation does not require that the systems have the same cardinality (as it is necessary for an isomorphism holding between two systems). In practice, it happens quite often that systems are associated via a bisimulation relation with different cardinalities of the associated sets X and X'.

With the above concepts we are able to introduce the anti-foundation axiom. This axiom requires the existence of a unique solution of every flat system of equations. Because we have defined four different types of systems it is possible to introduce four associated types of anti-foundation axioms. As it turns out the weakest form, namely the one that requires that every flat system of equations has a unique solution implies all stronger forms. We will come back to equivalent formulations later.

Definition 11.2.4 Anti-Foundation Axiom: *Every flat system of equations has a unique solution.*

In the remaining part of this section, we will consider examples that can clarify the situation. Important is to realize how the concepts system of equations, solution of a system, and the claim that there is a unique solution for an arbitrary system are connected with each other.

Example 11.2.2 (i) Let $\mathcal{E} = \langle X, A, e\rangle$ be given by: $X = \{x\}$, $A = \emptyset$, and $e : x \mapsto \{x\}$. If it holds $x \in \mathcal{U}$, we see immediately that $\mathcal{E}$ is a flat system of equations. A solution to $\mathcal{E}$ is a function s with $dom(s) = X = \{x\}$ and $s(x) = \{s(x)\}$. Applying the anti-foundation axiom, we know that such a set exists uniquely. Hence, it holds: $\{s(x)\}$ is the solution-set for $\mathcal{E}$. In standard works of non-well-founded set theory, $s(x)$ is usually denoted by Ω. Note: Ω does not exist in the well-founded universe of ZFC. In a certain sense, Ω is the simplest non-well-founded set (reflexive set) that can be constructed.

(ii) We can apply the anti-foundation axiom also to systems of equations which use only well-founded sets as atoms. For instance, if we wish to get as solution-set $\{2,3\} = \{\{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}\}$, we simply have to consider the system $\mathcal{E} = \langle X, A, e\rangle$ with $X = \{x, y\} \subseteq \mathcal{U}$, $A = \{0, 1, 2\}$ and a function $e : X \longrightarrow \wp(X \cup A)$, such that $e(x) = \{0, 1\}$ and $e(y) = \{0, 1, 2\}$.[12] It can easily be checked that the solution-set $\{s(x), s(y)\}$ for $\mathcal{E}$ gives us precisely the desired result $\{2, 3\}$.

(iii) Consider an arbitrary flat system $\mathcal{E} = \langle X, A, e\rangle$ with $X \subseteq \mathcal{U}$ and $A = \emptyset$. Which possible solutions do we get by varying e? Because $A = \emptyset$, one could think that we have only very restricted possibilities. But this is not true. For example, we can easily generate the natural numbers (considered as elements of the well-founded universe) as possible solutions of an appropriate system of equations. The following system represents precisely this idea.

$$\begin{aligned} e(x_1) &= \emptyset \\ e(x_2) &= \{x_1\} \\ e(x_3) &= \{x_1, x_2\} \\ \vdots \quad &\vdots \quad \vdots \end{aligned}$$

Clearly, the solution-set determines the collection of natural numbers. Furthermore, it is possible to form sets that contain Ω and arbitrary natural numbers: For example, consider the mapping $e(x') = \{x_0, x_1, x_3\}$ with $e(x_0) = \{x_0\}$. Hence, we are quite flexible, although we have no urelements at hand to form more complex sets.[13] One can show that even with $A = \emptyset$ we are able to construct the whole pure non-well-founded universe, where the well-founded sets (although already given by the axioms of ZFC) can be modeled as solutions of equations.

(iv) Every set can be an element of the solution set of an appropriate canonical system of equations. For instance, assume $a = \{c, \{c, d\}\}$ with $c \in \mathcal{U}$ and $d \in \mathcal{U}$ is given. The corresponding canonical system has the form

$$\mathcal{E} = \langle TC(\{a\}) - \mathit{support}(a),\ \mathit{support}(a),\ e\rangle$$

with the following first argument:

[12] The solution-set of a system $\mathcal{E} = \langle X, A, e\rangle$ is defined as follows:

$$\mathit{solution} - \mathit{set}(\mathcal{E}) = \{s(x) \mid x \in X\}$$

[13] Metaphorically speaking, we can say, every flat system has at least one 'kind of urelement' as atom: namely the empty set. (Certainly, the empty set is not literally an urelement.)

$$\begin{aligned} TC(\{a\}) - support(a) &= \{c, d, \{c, d\}, \{c, \{c, d\}\}\} - \{c, d\} \\ &= \{\{c, d\}, \{c, \{c, d\}\}\} \end{aligned}$$

Therefore it holds:

$$e(\{c, \{c, d\}\}) = \{c, \{c, d\}\} = a \ \in \ solution - set(\mathcal{E})$$

(v) We add some remarks with respect to collections that cannot be solutions of flat systems (or other versions of system of equations). We saw already that there is no set satisfying the equation $a = \wp(a)$. Other collections that are excluded are proper classes. Equations like $x = \mathcal{V}$ or $x = ORD$ are not allowed as appropriate systems.[14] Hence, there cannot be a solution for such equations. Further examples are equations of the form $x = x$ where $x \in \mathcal{U}$. Clearly, such equations cannot have unique solutions, because every set would satisfy this equation.

Now, we are able to examine the connections between different kinds of systems. The next section is devoted to this topic.

11.3 Connections between Systems

In the above section, we introduced a variety of different systems of equations. A natural question is which relations do hold between these types of systems. We want to specify these relations, in order to get a better understanding of these important concepts.

The following proposition specifies the connection between flat systems of equations and generalized flat systems.

Proposition 11.3.1 *The following assertions are equivalent:*

(i) Every flat system of equations has a unique solution.
(ii) Every generalized flat system of equations has a unique solution.

Proof. "(i) ⇒ (ii)": Assume that every flat system of equations has a unique solution. We have to prove that every generalized system $\mathcal{E} = \langle X, A, e\rangle$ has a unique solution. We introduce for each $x \in X$ a new urelement y defined by $y_x := \texttt{new}(x, A)$. Applying the two place functions `new` to every $x \in X$ (relative to a fixed A), we generate a set Y, such that $Y = \{y_x \mid x \in X\}$. A direct consequence of the definition of $\texttt{new}(x, A)$ are the following two relations: $Y \cap A = \emptyset$ and $Y \subseteq \mathcal{U}$.
Now we can define a system of equations $\mathcal{E}' = \langle Y, A, e'\rangle$ with the following specification of the function e':

[14] This can be easily checked by the definitions.

$$e'(y_x) = \{y_z \mid z \in e(x) \cap X\} \cup (e(x) \cap A)$$

It is easy to check that $\mathcal{E}'$ is a flat system of equations. Because of our assumption that every flat system has a unique solution, we can deduce that $\mathcal{E}'$ has a unique solution s' as well. In order to see this, define the solution of $\mathcal{E}$ by $s(x) = s'(y_x)$. s is a solution for $\mathcal{E}$, because the following equivalences hold:

$$\begin{aligned} s(x) = s'(y_x) &= \{s'(y_{x'}) \mid y_{x'} \in e'(y_x) \cap Y\} \cup (e'(y_x) \cap A) \\ \Leftrightarrow\ & s'(y_x) = \{s'(y_{x'}) \mid x' \in e(x) \cap X\} \cup (e(x) \cap A) \\ \Leftrightarrow\ & s(x) = \{s(x') \mid x' \in e(x) \cap X\} \cup (e(x) \cap A) \end{aligned}$$

It is important to notice that the form of s suffices to determine that s is a solution of $\mathcal{E}$. Because every solution of $\mathcal{E}$ is also a solution for $\mathcal{E}'$, $\mathcal{E}$ must have a unique solution, too.

"(ii) $\Rightarrow$ (i)": This direction is an obvious consequence of the fact that flat systems are a subclass of generalized flat systems. q.e.d.

We are able - using the anti-foundation axiom - to form sets which are not in the well-founded universe, provided these sets are solutions of an appropriate system of equations. Although we know the existence of certain sets, there is no real transparent method to compare different solutions of systems of equations. Whether a solution $s(x)$ of a system $\mathcal{E}$ is equal to a solution s' of a system $\mathcal{E}'$ cannot be determined in general. Notice that the axiom of extensionality, although a sufficient axiom for non-circular sets does not work in the theory of hypersets. In order to see this, consider the circular set $\Omega = \{\Omega\}$. Applying the axiom of extensionality we get the following infinite regress of equalities of sets without having a possibility to stop this process.

$$\Omega = \{\Omega\} = \{\{\Omega\}\} = \{\{\{\Omega\}\}\} = \ldots$$

Fortunately, there is a strong tool we can use, namely the concept of the bisimilarity of two sets. Using bisimulations instead of the axiom of extensionality enables us to compare non-well-founded sets. We postpone the introduction of bisimulations between sets to the next section. Here, we can use bisimulations between systems in order to have a possibility to compare systems of equations. The following theorem states that two systems of equations are bisimilar if and only if they have the same solution set.

Theorem 11.3.2 *Let $A \subseteq \mathcal{U}$ be given. Additionally, let $\mathcal{E} = \langle X, A, e\rangle$ and $\mathcal{E}' = \langle X', A', e'\rangle$ be two generalized flat systems of equations, with corresponding solutions s and s', respectively. Then it holds:*

$$\textit{solution-set}(\mathcal{E}) = \textit{solution-set}(\mathcal{E}') \quad \Leftrightarrow \quad \mathcal{E} \textit{ and } \mathcal{E}' \textit{ are bisimilar}$$

Proof: "$\Rightarrow$" Assume the solution-sets of $\mathcal{E}$ and $\mathcal{E}'$ are equal, i.e. solution-set($\mathcal{E}$) = solution-set($\mathcal{E}'$). We define a relation $R \subseteq X \times X'$ by the natural condition: xRx' iff $s(x) = s'(x')$. We have to show that R is a bisimulation relation, i.e. we need to check that the conditions specified in Definition 11.2.3 are satisfied.
First, let us consider condition (a) in Definition 11.2.3(i): Assume that $x \in X$ and $x' \in X'$, such that xRx' holds. Assume further that $y \in e(x) \cap X$ is arbitrarily given. Because $\mathcal{E}$ and $\mathcal{E}'$ have the same solution sets and $s(x) = s'(x')$, we can infer that for y there is a $y' \in e(x') \cap X'$, such that $s(y) = s'(y')$. Using the definition of R this means that it holds yRy'. Therefore, the first condition is satisfied.
A similar argument can be used to show that condition (b) of Definition 11.2.3(i) is satisfied.
In order to check condition (c) in Definition 11.2.3(i), we assume xRx' for $x \in X$ and $x' \in X'$. We know that it holds: $s(x) = s'(x')$, because of the definition of R. Therefore, $s(x) \cap A = s'(x') \cap A$. The following equalities do obviously hold:

$$\begin{aligned} s(x) \cap A &= [\{s(y) \mid y \in e(x) \cap X\} \cup (e(x) \cap A)] \cap A \\ &= [\{s(y) \mid y \in e(x) \cap X\} \cap A] \cup [(e(x) \cap A) \cap A] \\ &= \emptyset \cup (e(x) \cap A) \\ &= e(x) \cap A \end{aligned}$$

The equalities follow from ordinary set theory and the fact that $A \subseteq \mathcal{U}$. Similar equalities do hold for $s'(x')$ as well. We can conclude that it holds: $e(x) \cap A = s(x) \cap A = s'(x') \cap A = e'(x') \cap A$.
It remains to show that conditions (a) and (b) from Definition 11.2.3(ii) do hold. Let $x \in X$ be arbitrarily given. Because $\mathcal{E}$ is a flat system, $s(x) \in$ solution-set($\mathcal{E}$). The solution sets of $\mathcal{E}$ and $\mathcal{E}'$ are equal, hence there is a $x' \in \mathcal{E}'$, such that $s(x) = s'(x')$. Therefore, using our definition of R, it holds xRx'. A similar argument can be given to show condition (b) of Definition 11.2.3(ii). This shows one direction of the claim.

"$\Leftarrow$" We assume that $\mathcal{E}$ and $\mathcal{E}'$ are bisimilar guaranteed by a bisimulation relation R. First, we show that if xRx' holds, then $s(x)Rs'(x')$ does also hold. To prove this implication we introduce a new system of equations $\mathcal{E}^* = \langle X^*, A, e^* \rangle$, with the following properties:

$$\begin{aligned} &X^* = \{\langle x, x' \rangle \mid x \in X \wedge x' \in X' \wedge xRx'\} \\ &e^*(\langle u, u' \rangle) = \{\langle v, v' \rangle \mid v \in e(u) \cap X \wedge v' \in e'(u') \cap X'\} \cup (e(u) \cap A) \end{aligned}$$

Notice that in $\mathcal{E}^*$ there are the same atoms involved as in $\mathcal{E}$ and in $\mathcal{E}'$, respectively. Furthermore: the intersection of X^* and A is empty, and e^* is a mapping sending $\langle x, x' \rangle \in X^*$ into $\wp(X^* \cup A)$. Therefore, we can justify that $\mathcal{E}^*$ is a generalized flat system of equations. With the anti-foundation axiom and Proposition 11.3.1, there is a unique solution s^* of $\mathcal{E}^*$. Because s^* is a solution of $\mathcal{E}^*$, solution s^* has necessarily the following form:

$$\begin{aligned} s^*(\langle u,u'\rangle) &= \{s^*(\langle v,v'\rangle) \mid \langle v,v'\rangle \in e^*(\langle u,u'\rangle) \cap X^*\} \cup (e^*(\langle u,u'\rangle) \cap A) \\ &= \{s^*(\langle v,v'\rangle) \mid v \in e(u) \cap X \wedge v' \in e'(u') \cap X'\} \cup (e(u) \cap A) \end{aligned}$$

Now, define $s_1(\langle u,u'\rangle) = s(u)$ and $s_2(\langle u,u'\rangle) = s'(u')$, respectively. We have to prove that s and s' are in fact solutions for $\mathcal{E}^*$. In order to prove this claim, we need to show that the following formula holds for s_1 (and also for s_2):

$$(\dagger) \quad s_1(\langle u,u'\rangle) = \{s_1(\langle v,v'\rangle) \mid \langle v,v'\rangle \in e^*(\langle u,u'\rangle) \cap X^*\} \cup (e^*(\langle u,u'\rangle) \cap A)$$

In order to show (†), we need to show that the following equivalence holds.

$$\begin{gathered} b \in s_1(\langle u,u'\rangle) = s(u) \\ \Leftrightarrow \; b \in \{s_1(\langle v,v'\rangle) \mid \langle v,v'\rangle \in e^*(\langle u,u'\rangle) \cap X^*\} \cup (e^*(\langle u,u'\rangle) \cap A) \end{gathered}$$

If $b \in s(u)$, then b is either of the form $s(v)$ for $v \in e(u) \cap X$, or $b \in e(u) \cap A$. If $b = s(v)$, then there exists $v' \in e'(u') \cap X'$, such that vRv', therefore $\langle v,v'\rangle \in X^*$. Additionally, we know: in accordance to the definitions above, we get the relation $\langle v,v'\rangle \in e^*(\langle u,u'\rangle) \cap X^*$. Hence, we have:

$$s_1(\langle v,v'\rangle) = b \in \{s_1(\langle v,v'\rangle) \mid \langle v,v'\rangle \in e^*(\langle u,u'\rangle) \cap X^*\} \cup (e^*(\langle u,u'\rangle) \cap A)$$

In the second case, if $b \in e(u) \cap A$, then it holds obviously $b \in e^*(\langle u,u'\rangle) \cap A$.

To show the converse, we assume that it holds:

$$b \in \{s_1(\langle v,v'\rangle) \mid \langle v,v'\rangle \in e^*(\langle u,u'\rangle) \cap X^*\} \cup (e^*(\langle u,u'\rangle) \cap A)$$

If $b = s_1(\langle v,v'\rangle)$ for $v \in e(u) \cap X$ and $v' \in e'(u') \cap X'$, then $b = s(v) \in s(u)$, because s is a solution of $\mathcal{E}$. On the other hand, if $b \in e^*(\langle u,u'\rangle) \cap A$, then $b \in e(u) \cap A$, by definition. Therefore: s is in fact a solution of $\mathcal{E}^*$.

To see that s_2 is a solution of $\mathcal{E}^*$, we need only to consider the dual case. Notice that the apparent asymmetric condition $e(u) \cap A$ in the definition of $e^*(\langle u,u'\rangle)$ does not cause any serious problems: because $\mathcal{E}$ and $\mathcal{E}'$ are bisimilar, we know that it holds by definition of bisimulations between systems: $e(u) \cap A = e'(u') \cap A$ for u, u' with uRu'.

We have proven that if x and x' are bisimilar, then $s(x) = s'(x')$. It remains to show, that the solution-sets of $\mathcal{E}$ and $\mathcal{E}'$ are equal. Take an arbitrary set $b = s(x) \in$ solution-set$(\mathcal{E})$ for $x \in X$. We know that there is a bisimulation between $\mathcal{E}$ and $\mathcal{E}'$, therefore there is an $x' \in X'$ with xRx', and because the implication $xRx' \rightarrow s(x) = s'(x')$ does hold, we can justify the equivalence $s(x) = s'(x')$. Conclude: $s(x) = b \in$ solution-set$(\mathcal{E}')$. The converse argument holds similarly. Conclude: solution-set$(\mathcal{E})$ = solution-set$(\mathcal{E}')$. q.e.d.

Remark 11.3.1 (i) Proposition 11.3.1 and Theorem 11.3.2 make clear that the anti-foundation axiom can be stated in different but equivalent forms. If we assume that every flat system of equations has a unique solution, then this implies that every generalized flat system has a unique solution, and vice versa. In fact, both statements are equivalent. That implies that the anti-foundation axiom can be introduced at least in two different forms. The comparability of the two different types of systems is guaranteed by the bisimulation relation between the systems.

(ii) Whereas, two systems can appear to be quite different they, nevertheless, can have the same solution-sets. This insight emphasizes the structural similarity guaranteed by a bisimulation relation.

Example 11.3.2 We give an example of two systems $\mathcal{E} = \langle X, A, e\rangle$ and $\mathcal{E}' = \langle X', A', e'\rangle$ that are bisimilar. Assume $\mathcal{E}$ is given by $X = \{x_1, x_2\}$, $A = \emptyset$, and $e : X \longrightarrow \wp(X \cup A)$ is specified by

$$\begin{aligned} &e(x) = \{y\} \\ &e(y) = \{x\} \end{aligned}$$

Intuitively, $\mathcal{E}$ is similar to a truth-teller expression. The system $\mathcal{E}'$ is specified by $X' = \{x'\}$, $A = \emptyset$, and $e : X' \longrightarrow \wp(X' \cup A)$, such that

$$e(x') = \{x'\}$$

It is easy to see that the two systems are bisimilar. Furthermore, an easy calculation shows that solution-set($\mathcal{E}$) = solution-set($\mathcal{E}'$). Notice that the two systems are not isomorphic, because they have different cardinalities. Using bisimulations we are able to associate truth-teller-like expressions with truth-teller-like circles.[15]

In the next subsection, we want to examine the concept of bisimulations more closely. In particular, we are interested to use bisimulations to compare sets (not only systems of equations).

11.4 Bisimulations on Sets

We saw above that for non-well-founded sets it is not possible to apply the axiom of extensionality of *ZFC* set theory in general. Comparing two sets using the axiom of extensionality yields an infinite regress. The ordinary axiom of extensionality provides no effective tool to compare reflexive sets. We can use the (ordinary) axiom of extensionality to decide the same question for well-founded sets, but in case of two non-well-founded sets, the question is not

[15] Clearly the examples are literally not truth-teller expressions. We are working on a purely set theoretic level, without a truth predicate and a specified formal language (besides the language of set theory).

straightforwardly decidable. We give a further example. Consider the sets $a = \{b, u\}$ and $b = \{a, u\}$ where u is an urelement. The axiom of extensionality tells us that $a = b$ iff a and b have the same elements, but this means nothing else, then $a = b$ iff $u = u$ and $b = a$. Whereas $u = u$ is trivially satisfied, the problem remains whether it holds $b = a$ or not. Since the latter equality states simply the initial problem again it is not very informative. We will formulate a condition in Theorem 11.4.2 which gives us a useful tool to compare non-well-founded sets. Before we can do this, we need to define the notion of a bisimulation on sets (which is strongly related to the notion of a bisimulation between systems of equations.)

Definition 11.4.1 *A bisimulation R on two sets a and b is a binary relation $R \subseteq a \times b$, such that the following conditions (i)-(iii) hold:*

(i) $aRb \rightarrow (\forall a' \in a)(\exists b' \in b) : a'Rb'$
(ii) $aRb \rightarrow (\forall b' \in b)(\exists a' \in a) : a'Rb'$
(iii) $aRb \rightarrow a \cap \mathcal{U} = b \cap \mathcal{U}$

Notice the obvious similarity of Definition 11.4.1 and Definition 11.2.3. On the other hand there are also differences: the restriction of bisimulations on sets is crucial for our development. Whereas a bisimulation between systems can be a binary relation defined on urelements, this is not allowed for bisimulations on sets. Intuitively, two sets a and b are bisimilar if and only if every element of a is bisimilar to a certain element of b and vice versa.

Using Definition 11.4.1 we can state the crucial relation between two non-well-founded sets in order to guarantee the equality of these sets. To say that two sets bisimulate is nothing else than to say that the sets are extensionally equal. This is the claim of the next theorem.

Theorem 11.4.2 *Two sets are equal if and only if they are bisimilar.*

Proof: "$\Rightarrow$" Consider the equality relation $=$ on sets. We need to show that this relation $=$ is a bisimulation. This is obvious: for instance, assume for two sets a, b that it holds $a = b$, then trivially the condition $(\forall a' \in a)(\exists b' \in b) : a' = b'$ holds, because a and b have precisely the same elements. The other conditions are shown similarly.

"$\Leftarrow$" Assume that R is a bisimulation relation on sets. We have to prove that, if aRb for two sets a and b, then it holds also $a = b$. In order to show this, we prove that R is a subrelation of the equality relation on sets. Construct the canonical systems of equations $\mathcal{E}$ and $\mathcal{E}'$ (relative to given sets a and b, respectively) that are specified as follows:

$$\mathcal{E} = \langle X, A, e \rangle = \langle TC(\{a\}) - support(a),\ support(a),\ e \rangle$$
$$\mathcal{E}' = \langle X', A', e' \rangle = \langle TC(\{b\}) - support(b),\ support(b),\ e' \rangle$$

First, we prove that $a = A'$. Consider an arbitrary $u \in A = \mathit{support}(a)$. By the definition of the support operation, there is an a', such that $u \in a' \in TC(\{a\})$. By the definition of the transitive closure on sets, it holds: there are sets $a_1, a_2, \dots a_n$, such that $a' \in a_n \in a_{n-1} \in \dots \in a_1 \in a$. We can use condition (i) in Definition 11.4.1 that there is $b' \in \mathit{support}(b) = A'$, such that $a'Rb'$. The bisimulation relation R holds also for all a_is. We summarize these considerations in the following formula:

$$\begin{aligned} (\forall a_i)(\exists b_i): \quad & (i \in \{1, 2, \dots, n\} \;\wedge\; (a' \in a_n \in a_{n-1} \in \dots \in a_1 \in a) \\ \wedge \quad & (b' \in b_n \in b_{n-1} \in \dots \in b_1 \in b)) \;\longrightarrow\; a_iRb_i) \end{aligned}$$

The above fact can be interpreted as follows: because aRb holds by assumption, we know that for $a_1 \in a$ there is $b_1 \in b$, such that a_1Rb_1. Furthermore, because a_1Rb_1 it holds: for every $a_2 \in a_1$ there exists $b_2 \in b_1$, such that a_2Rb_2. Hence, we apply condition (i) from Definition 11.4.1 $n+1$ many times to get $a'Rb'$. By the definition of the transitive closure of a set and the definition of a bisimulation on sets, we know that $b' \in TC(\{b\})$ and $u \in \mathit{support}(b)$, and therefore $A \subseteq A'$. The same argument from the beginning works also for the converse inclusion. Conclude: $A = A'$.

Now we show that the restriction $R^* = R \restriction X \times X'$ is a bisimulation between $\mathcal{E}$ and $\mathcal{E}'$. First, we check condition (i) of Definition 11.4.1 by constructing the set $Y = \{x \mid x \in X \wedge \exists x' \in X' : xR^*x'\}$ and prove that $X = Y$. Obviously $Y \subseteq X$. To show the converse notice first that $a \in Y$ (because a and b bisimulate), and second that for $y \in x \in Y$, it holds $y \in Y$. To see the second claim, notice that $xRx' \rightarrow (\forall y' \in x')(\exists y \in x) : yRy'$. Assume $y \notin Y$. Then $y' \in X'$ (because X' is transitive) and $y \in X$, hence $y \in Y$. We have a contradiction. We have shown that Y is a transitive set including a. Because $TC(\{a\}) - \mathit{support}(a)$ is the smallest transitive set including a, we can deduce: $TC(\{a\}) - \mathit{support}(a) \subseteq Y$. Together we get: $X = Y$. Condition (ii) of Definition 11.4.1 can by shown by a similar argument.

It remains to show condition (i) of Definition 11.4.1: Assume xR^*y and $x' \in e(x) \cap X$. Because e is the identity function mapping X into X, we have $x' \in x \cap X$, and therefore x' is a set. Using the bisimulation R on sets, there is some $y' \in y$, such that $x'Ry'$, and with the transitivity of X' we get: $y' \in X'$. Conclude: $y' \in e'(y) \cap X'$ with the property $x'R^*y'$.

The other conditions can be checked in a similar way. Therefore R^* is in fact a bisimulation between $\mathcal{E}$ and $\mathcal{E}'$. Using this fact and Theorem 11.3.2, we can state the following equivalences:

$$a = s(a) = s'(b) = b$$

This suffices to show the theorem. q.e.d.

We mention some easy examples where one can see immediately that the application of bisimulations simplify matters significantly.

Example 11.4.1 (i) Consider the reflexive set Ω that satisfies the equation $\Omega = \{\Omega\}$. It is easy to see that there is a bisimulation between Ω and $\{\Omega\}$ specified by $R = \{\langle\{\Omega\},\Omega\rangle, \langle\Omega,\{\Omega\}\rangle, \langle\Omega,\Omega\rangle, \langle\{\Omega\},\{\Omega\}\rangle\}$. By our definition Ω and $\{\Omega\}$ are also equal, i.e. they denote the same set theoretical entity.

(ii) Consider two sets a and b specified as follows.

$$a = \{a, b\}$$
$$b = \{a, b\}$$

Then $R = \{\langle b,a\rangle, \langle a,b\rangle, \langle a,a\rangle, \langle b,b\rangle\}$ is a bisimulation. Clearly, it holds also: $a = b$.

Remark 11.4.2 We will see in the following chapters that other structures require slightly different definitions of bisimulations. In particular, the property that every bisimulation is an equivalence relation is not generally true in every framework. For the theory of hypersets this property simplifies many things, because bisimulations can be used in order to test whether two given sets are equal or not.

In the following section, we will state some properties of the non-well-founded universe. Interestingly enough, many properties of 'classical' concepts are preserved in the context of non-well-founded set theory, although they are based seemingly on a well-founded universe.

11.5 The Non-Well-Founded Universe $V_{afa}[\mathcal{U}]$

We will summarize some further properties of hypersets that are interesting from a theoretical point of view. Although the facts presented in this section are not important for the development of the theory of non-well-founded sets in the following parts of this work, they are illuminating and interesting. Proofs of the corresponding claims can be found in [BarMo96].

First, we will consider some properties of the non-well-founded universe $V_{afa}[\mathcal{U}]$. This universe is specified by the axioms of ZFC_{afa}. These axioms are essentially the axioms of ZFC^- plus axioms that guarantee that there is a class of urelements and that every flat system of equations has a unique solution.[16]

Essentially, there are two possibilities to define the universe of hypersets. One possibility is to collect all sets that are compatible with the axioms of the theory of hypersets ZFC_{afa}. The second possibility is to collect all solution-sets of systems of equations. Hence, there are the following two equivalent characterizations of the universe (relative to a collection $X \subseteq \mathcal{U}$ of urelements):

[16] ZFC^- stands for the axioms of ZFC without the foundation axiom.

Fact 11.5.1 *Assume the axioms of ZFC_{afa} are given. Then, the universe $V_{afa}[X]$ can be equivalently specified by (a) and (b).*

$$\begin{array}{lll} (a) & V_{afa}[X] = & \{a \mid a \textit{ is a set with support}(a) \subseteq X\} \\ (b) & V_{afa}[X] = & \{\textit{solution-set}(\mathcal{E}) \mid \mathcal{E} \textit{ is a flat system of equations} \\ & & \quad \textit{with atoms } A \subseteq X\}] \end{array}$$

Proof: Compare [BarMo96]. q.e.d.

As we can see from this characterization, we can identify $V_{afa}[X]$ with the collection of all solutions of all flat systems using elements of X as atoms. On the other hand we can identify $V_{afa}[X]$ with the collection of all sets which can be built by the axioms of set theory using elements of X as urelements. From characterization (a) it follows that every set in the classical ZFC universe is also in the non-well-founded universe. This follows from the fact that we can build the same sets as in ZFC, but additionally we can build sets that are circular, i.e. sets that are not prohibited by the foundation axiom.

There are a number of results of properties of the non-well-founded universe. We mention some of these properties. The first fact shows that the ordinary set theoretical definition of the successor function is appropriate even for the non-well-founded universe.

Fact 11.5.2 *In $V_{afa}[X]$, the set theoretical successor function is injective.*

Proof: Compare [BarMo96]. q.e.d.

Fact 11.5.2 is a surprising result. The claim is that even for non-well-founded sets a and b it holds:[17]

$$s(a) = s(b) \;\rightarrow\; a = b$$

This property is clearly satisfied in the case of well-founded sets. For hypersets this is less clear.[18] We consider an example.

Example 11.5.1 Consider the two sets $s(a)$ and $s(b)$ specified as follows:

$$\begin{array}{lll} s(a) & = & \{\Omega, \emptyset\} \cup \{\{\Omega, \emptyset\}\} \;=\; \{\Omega, \emptyset, \{\Omega, \emptyset\}\} \\ s(b) & = & \{\{\Omega\}, \emptyset\} \cup \{\{\{\Omega\}, \emptyset\}\} \;=\; \{\{\Omega\}, \emptyset, \{\{\Omega\}, \emptyset\}\} \end{array}$$

It is easy to see that there exists a bisimulation between $s(a)$ and $s(b)$, hence they are equal. From the fact that the successor function is injective, we can infer that $a = \{\Omega, \emptyset\}$ and $b = \{\{\Omega\}, \emptyset\}$ are also equal.

[17] $s(a)$ denotes the successor of a.

[18] As a consequence of this fact, the axiom of infinity in which we used a certain notion of finiteness in terms of the successor function is literally the same as in ZFC.

Using the injectivity of the successor function, we can characterize reflexive sets. The following fact makes this precise.

Fact 11.5.3 *In the non-well-founded universe, the following properties of a set a are equivalent:*

(a) $a \in a$
(b) $s(a) = a \cup \{a\} = a$
(c) $s(a) \in a$

Proof: Compare [BarMo96]. q.e.d.

It could seem that in $V_{afa}[X]$ nearly every conceivable collection is included. But this is not the case. We know already that we can specify a condition that does not determine a set (collection): Even in the non-well-founded universe there is no collection a that satisfies the equation $a = \wp(a)$ (by Cantor's theorem). Another example is the one-element collection $\{C\}$ where C is a proper class. In non-well-founded set theory, $\{C\}$ is a proper class and not a set. Notice also that every proper class, i.e. a collection that is not in the universe of ZFC (for example the class of all ordinals, the class of all sets, the class of all singletons etc.) is not an element of V_{afa}, either. In other words, the universe of non-well-founded set theory is a restricted extension of the universe of ZFC.

We can use a form of a diagonal argument to show that another collection is not an element of $V_{afa}[\mathcal{U}]$. Assume a is a transitive set. Then, there is no set $b = a - \{a\}$ which satisfies $b \in a$. Clearly, there is no hope that such a set b exists in the well-founded universe. Using hypersets, it is on first sight a strange result that there should not be a set b which collects all elements of a that are not equal to a and simultaneously is an element of a.

In the following fact, we summarize the above considerations.

Fact 11.5.4 *Consider the universe $V_{afa}[\mathcal{U}]$. Then, the following statements hold:*

(a) There is no set $a \in V_{afa}[\mathcal{U}]$, such that $a = \wp(a)$
(b) If C is a class, then $\{C\} \notin V_{afa}[\mathcal{U}]$
(c) If a is transitive, then it holds:

$$\neg\exists b : b = a - \{a\} \ \wedge \ b \in a$$

Proof: Compare [BarMo96]. q.e.d.

We mention a last remark. Because we work with urelements some sets in our universe contain elements that are themselves non-sets. An example is the set $a = \{u\}$ where $u \in \mathcal{U}$. As a consequence it follows that every urelement

$x \in \mathcal{U}$ is not an element of $V_{afa}[U]$, because x is not a set and $V_{afa}[U]$ is a collection of sets.

We have to be very careful which collections can count as subsets of $V_{afa}[\mathcal{U}]$. Notice that for $x \in \mathcal{U}$ it holds $\{x\} \in V_{afa}[\mathcal{U}]$, but $\{x\} \not\subseteq V_{afa}[\mathcal{U}]$. The second claim follows from the fact that for $x \in \mathcal{U}$ it does not hold: $x \in \{x\} \rightarrow x \in V_{afa}[\mathcal{U}]$. This phenomenon can be generalized to every flat set of urelements: for $x_1, x_2, \dots \in \mathcal{U}$ it holds that $\{x_1, x_2, \dots\} \not\subseteq V_{afa}[\mathcal{U}]$. Nevertheless we have $\{x_1, x_2, \dots\} \in V_{afa}[\mathcal{U}]$. Notice that these properties are not a consequence of the anti-foundation axiom, but a consequence of the usage of urelements.

In the next section, we will consider general systems of equations and a very important concept, namely the concept of corecursive substitution. This principle is constitutive for the whole theory of hypersets.

11.6 General Systems and Corecursive Substitution

In Definition 11.2.1, we introduced flat, generalized, and general systems of equations. In this section, we will examine the properties of the last kind of systems of equations. Notice that we are able to solve systems of equations $\mathcal{E} = \langle X, A, e\rangle$ where $e : X \longrightarrow \wp(X \cup A)$ using the anti-foundation axiom. Now, consider the system $\mathcal{E}' = \langle X', A', e'\rangle$ where $X' = \{x\}$, $A' = \{u\}$, and $e'(x) = \{u, \{x, u\}\}$. It is clear that we cannot apply (straightforwardly) the anti-foundation axiom, because the set $\{u, \{x, u\}\} \notin \wp(\{x, u\}) = \wp(X' \cup A')$. An intuitively appropriate solution for $\mathcal{E}'$ is the following flat system $\mathcal{E}''$ (resulting from $\mathcal{E}'$ by substituting y for $\{x, u\}$): $\mathcal{E}'' = \langle X'', A', e''\rangle$ with $X'' = \{x, y\}$, A' as above, and

$$e''(x) = \{u, y\}$$

$$e''(y) = \{x, u\}$$

What we did is to use a substitution in order to transform the given system $\mathcal{E}'$ into a flat system $\mathcal{E}''$. (At the same time the number of indeterminates increases.) The solution-set of $\mathcal{E}''$ is given by:

$$\text{solution-set}(\mathcal{E}'') = \{\{s(x), u\}, \{u, s(y)\}\}$$

At first sight, it seems that we simply have to apply ordinary substitution to obtain the desired result. Unfortunately, ordinary substitution does not work, because for non-well-founded sets no appropriate substitution operation is defined. First, we need to introduce a notion of substitution which is well-defined in the framework of non-well-founded sets. Whether this is possible at all, is less clear, because we are not able to specify a base case in the substitution extending this base case step by step in n many (or α many) inductive steps. We give an example that shows the problems of a substitution operation for non-well-founded sets.

Example 11.6.1 Assume a set $x = \{u, \{x, u\}\}$ is given. Assume further that we want to define a substitution operation defined on x substituting u by the set $4 = \{0, 1, 2, 3\}$. Then, the intuitively correct result of (ordinary) substitution yields $sub(x) = \{4, \{x, 4\}\}$. This specification of $sub(x)$ does not give us further information about x. Even if the substitution is performed we do not know whether it holds $\{4, \{\{4, \{x, 4\}\}, 4\}\} = \{4, \{x, 4\}\}$ or not. There is a substitution instance for the embedded x missing. Therefore, we must ensure to give a definition of substitution which takes care of the non-well-foundedness of our equations. Furthermore, we would like to find a substitution operation that allows to substitute sets that are deeply embedded in another set.

We will define a corecursive substitution operation and we will show that this substitution operation exists, is well-defined, and unique. Last but not least, we will prove that every general system of equations has a unique solution provided that the anti-foundation axiom holds.

Definition 11.6.1 *(i) A substitution s is a function with the properties: $dom(s) = X \subseteq \mathcal{U}$ and $range(s) = Y \subseteq \mathcal{U} \cup V_{afa}[\mathcal{U}]$.*

(ii) A substitution operation $sub(s, x)$ is an operation which is defined on the set of all pairs $\langle s, x\rangle$ where s is a substitution and $x \in \mathcal{U} \cup V_{afa}[\mathcal{U}]$, such that the following condition hold:

$$sub(s,x) = \begin{cases} s(x) & : \quad x \in \mathcal{U} \cap dom(s) \\ x & : \quad x \in \mathcal{U} - dom(s) \\ \{sub(s,x') \mid x' \in x\} & : \quad x \text{ is a set} \end{cases}$$

The conditions specified in Definition 11.6.1 are called corecursion conditions. Notice that crucial differences can arise between an ordinary well-founded substitution operation and a corecursive substitution operation because of the impact of the last condition. Whereas in every well-founded set we need to repeat the recursion process at most α many times (for an appropriate ordinal α), in the case of non-well-founded sets this is no longer true. For the set $\Omega = \{\Omega\}$ there is no ordinal α such that the corecursion defined above terminates. Because of that fact we need to know that Definition 11.6.1 is well-defined and determines a unique substitution operation. We examine an example in order to clarify the idea of Definition 11.6.1.

Example 11.6.2 Assume a set a of the following form is given.

$$a = \{\emptyset, x, \Omega, \{a, y\}\}$$

In a, the elements x and y are urelements. Although it does not hold $a \in a$, the set a has similarities to a reflexive set. Assume a substitution s is given as follows:

$$\{x, y\} \longrightarrow \{x, y\} \cup V_{afa}[\mathcal{U}]$$

We define s by the following conditions: $s(x) = 3$ and $s(y) = \Omega$. How does the non-well-founded substitution operation $sub(s, a)$ work on a? Clearly, the substitution operation satisfies $sub(s, x) = 3$, and $sub(s, \Omega) = \Omega$. How should the expression $sub(s, a)$ be interpreted? According to the definition of the corecursive substitution operation the following formula should be satisfied:

$$sub(s, a) = \{\emptyset, 3, \omega, \{sub(s, a), \{\Omega\}\}\}$$

Notice that the corecursive substitution operates on sets as the solution function in flat systems of equations. With this substitution we are able to go into the deep structure of a given set, without the necessity of unfolding the set provided corecursive substitution is well-defined.

The next theorem shows that the above substitution operation is in fact well-defined and unique, even if non-well-founded sets are involved.

Theorem 11.6.2 *The operation $sub(s, b)$ is well-defined on all pairs $\langle s, b\rangle$ where s is a substitution, $b \in \mathcal{U} \cup V_{afa}[\mathcal{U}]$, and additionally $sub(s, b)$ satisfies all corecursion conditions of Definition 11.6.1(ii). Furthermore, $sub(s, b)$ is uniquely defined.*

Proof: We need to show the existence of $sub(s, b)$ and the uniqueness of this set. Assume b is an arbitrary set and s is a substitution. Consider $sub(s, b)$ as defined in Definition 11.6.1(ii). The idea is to flatten b, such that the anti-foundation axiom can be applied to an appropriate system of equations. We consider $TC(b) - \mathcal{U}$. Introduce for every element $c \in TC(b) - \mathcal{U}$ a new urelement x_c. Define $X = \{x_c \mid c \in TC(b) - \mathcal{U}\}$. Enumerate all x_c appropriately. Then, we define the generalized flat system of equations $\mathcal{E} = \langle Y, A, e\rangle$ as follows: $Y = X \cup \{b\}$, $A = support(b)$, and e is specified by the following condition:

$$e_{y_\alpha} = \begin{cases} b & : \quad y_\alpha = b \\ s(a) & : \quad \text{the } \alpha^{th} \text{ element of } TC(b) - A \text{ is } a \wedge a \neq b \wedge a \in \wp(A) \\ y_\alpha \in Y : & \quad \text{the } \alpha^{th} \text{ element of } TC(b) - A \text{ is } a \wedge a \neq b \wedge a \notin \wp(A) \end{cases}$$

Clearly, $\mathcal{E}$ is a generalized (flat) system of equations. Hence, we can apply the anti-foundation axiom (using Fact 11.3.1). Consider the solution sol of $\mathcal{E}$. We need to check that sol satisfies the corecursion conditions of Definition 11.6.1(i) and (ii). Obviously, s is defined on urelements as is required in Definition 11.6.1. We need to verify the condition of sol on sets. Hence, the only interesting case is $sub(s, b)$ itself. For $sub(s, b)$ the following equalities hold:

$$\begin{aligned}
sub(s,b) &= sol(b) \\
&= \{sol(x) \mid x \in A \wedge x \in dom(s) \wedge x \in b\} \\
&\quad \cup \{sol(x) \mid x \in A \wedge x \notin dom(s) \wedge x \in b\} \\
&\quad \cup \{sol(y_\alpha) \mid y_\alpha \in Y\} \\
&= \{s_x \mid x \in A \wedge x \in dom(s) \wedge x \in b\} \\
&\quad \cup \{x \mid x \in A \wedge x \notin dom(s) \wedge x \in b\} \\
&\quad \cup \{s_{y_\alpha} \mid y_\alpha \in Y\} \\
&= \{sol(p) \mid p \in b\} \\
&= \{sub(s,p) \mid p \in b\}
\end{aligned}$$

We can conclude: $sol(s,b)$ satisfies literally the condition on sets of Definition 11.6.1.

In order to show uniqueness, we reason as follows. Assume there exists $sub'(s,b)$ that also satisfies the corecursion conditions of Definition 11.6.1. Consider the following relation.

$$R = \{\langle sub(s,b), sub'(s,b)\rangle \mid b \in V_{afa}[\mathcal{U}] \cup \mathcal{U}\}$$

Then, R is a bisimulation relation on sets. Clearly, the bisimulation condition for urelements are satisfied. For sets we have: if $p \in b$ is a set, then by definition the pair $\langle sub(s,p), sub'(s,p)\rangle \in R$. If $p \in b$ is an urelement, then by the definition of s it holds $sub(s,p) = sub'(s,p)$. Hence, $sub(s,b) = sub'(s,b)$. This suffices to show the theorem. q.e.d.

We add some remarks concerning the last theorem.

Remark 11.6.3 (i) The corecursive substitution is the basis on which the further development of the theory of non-well-founded sets is based. It follows from Theorem 11.6.2 that the anti-foundation axiom is equivalent to the axiom that there exists a unique corecursive substitution operation satisfying the conditions of Definition 11.6.1.

(ii) We will see later that there is a possibility to generalize the presented technique. In Chapter 14, we will develop a coalgebraic flattening procedure that generalizes the flattening of systems of equations using coalgebraic techniques.

(iii) By using corecursive substitution operations we are enabled to consider general systems and their solutions (justified by Theorem 11.6.3 below). It should be emphasized that substitution operations are quite important for certain applications, as for example in situation theory where the existence and uniqueness of solutions of general systems is crucially used.

We presented the precise definition of general systems in Definition 11.2.1(iv). Furthermore, in Definition 11.2.2(iv) we introduced the concept of a solution of a general system of equations. The step by step generalization from flat systems to generalized flat systems and finally general systems of equations culminates in the fact that every general system of equations has a unique solution (provided the anti-foundation axiom holds). This implies that the expressive power of the anti-foundation axiom suffices to ensure that general systems do have unique solutions. In other words, we do not strengthen the theory of hypersets if we define the anti-foundation axiom as follows: every general system has a unique solution. This is the crucial upshot of this section. It does not matter whether one speaks about flat systems, generalized flat systems, or general systems, because from a global perspective their generating power is equal. The following theorem formulates this precisely.

Theorem 11.6.3 *Every general system of equations $\mathcal{E} = \langle X, A, e\rangle$ has a unique solution s provided the anti-foundation axiom holds.*

Proof: We prove the assertion in two steps. First, we show that, if $\mathcal{E}$ has a solution s, then there exists a generalized flat system $\mathcal{E}' = \langle X', A, e'\rangle$, such that it holds:

(i) $\mathcal{E}'$ extends $\mathcal{E}$
(ii) The solution s' for $\mathcal{E}'$ extends s

As a second step we will prove that if s' is a solution for $\mathcal{E}'$ (the extension of $\mathcal{E}$), then the restriction $s \restriction X$ is a solution for $\mathcal{E}$. Then, existence and uniqueness of s are immediate consequences of these claims.
Define a generalized flat system $\mathcal{E}'$ as follows:

$$\mathcal{E}' = \langle [X \cup \bigcup_{x\in X} (TC(e(x)))] - A, A, e'\rangle$$

where e' is defined as follows:

$$e'(x) = \begin{cases} e(x) & : \quad x \in X \\ x & : \quad x \in X' - X \end{cases}$$

Notice: X contains only urelements which are also included in X'. First, we have to check that e' is appropriate, i.e. that e' is a function mapping X' into $\wp(X' \cup A) = \wp([X \cup \bigcup_{x\in X}(TC(e(x)))] - A)$.

Assume $x \in X$, then $e'(x) = e(x)$. Clearly, it holds $e(x) \in TC(e(x))$ by the definition of the transitive closure. Hence, $e(x) \subseteq \bigcup_{x\in X}(TC(e(x))$, and $e(x) \subseteq [X \cup \bigcup_{x\in X}(TC(e(x)))] - A$. That is why we can conclude that $e'(x) = e(x) \in \wp(X' \cup A)$. On the other hand, for $x \in X' - X$ it holds trivially that $e'(x) \in \wp(X' \cup A)$. Therefore: $e : X' \longrightarrow \wp(X' \cup A)$ and $\mathcal{E}'$ is a generalized flat system. Using the anti-foundation axiom and Proposition 11.3.1 we know that $\mathcal{E}'$ has a unique solution.

Now we show the following: Assume s is a solution of $\mathcal{E}$, then s extends to a solution s' of $\mathcal{E}'$ by the following specification of s':

$$\forall x \in X' : s'(x) := sub(s, x)$$

We prove that s' is in fact a solution of $\mathcal{E}'$ (notice: because $\mathcal{E}'$ is a generalized flat system, s' must be unique). In order to prove that s' is a solution, we need only to show that s' has the correct form.
Assume $x \in X' - X$. Then the following equalities hold:

$$\begin{array}{rcll}
s'(x) & = & sub(s,x) & \text{(Definition of } s'\text{)} \\
& = & \{sub(s,y) \mid y \in x\} & \text{(Defintion of } sub(s,y)\text{)} \\
& = & \{sub(s,y) \mid y \in x - A\} \cup (x \cap A) & \text{Logic} \\
& = & \{s'(y) \mid y \in x - A\} \cup (x \cap A) & \text{(Definition of } s'\text{)} \\
& = & \{s'(y) \mid y \in e'(x) - A\} \cup (x \cap A) & \text{(because } e'(x) = x\text{)} \\
& = & \{s'(y) \mid y \in e'(x) \cap X'\} \cup (e'(x) \cap A) & \text{(Definition of } X'\text{)}
\end{array}$$

Now we assume that $x \in X$. Then we can reason as follows:

$$\begin{array}{rcll}
s'(x) & = & s(x) & \text{(Definition of } s'\text{)} \\
& = & sub(s, e(x)) & \text{(Definition of } s\text{)} \\
& = & \{sub(s,y) \mid y \in e(x)\} & \text{(Definition of } sub(s,y)\text{)} \\
& = & \multicolumn{2}{l}{\{sub(s,y) \mid y \in e(x) \cap (X' - X)\} \cup \{sub(s,y) \mid y \in e(x) \cap X\}} \\
& \cup & \{sub(s,y) \mid y \in e(x) \cap A\} & \text{(Splitting of the elements of } e(x)\text{)} \\
& = & \multicolumn{2}{l}{\{s'(y) \mid y \in e'(x) \cap (X' - X)\} \cup \{s'(y) \mid y \in e'(x) \cap X\}} \\
& \cup & (e'(x) \cap A) & (e(x) = e'(x) \text{ and definition of } s'\text{)} \\
& = & \{s'(y) \mid y \in e' \cap X'\} \cup (e'(x) \cap A) & \text{(Logic)}
\end{array}$$

We can conclude: $s' = sub(s, x)$ is in fact a solution of $\mathcal{E}'$. Additionally, s' extends s.

The second step in the proof of the theorem consists in showing the following claim: if s' is the solution of $\mathcal{E}'$, then $s := s' \restriction X$ is a solution for $\mathcal{E}$. Assume s is a solution of $\mathcal{E}$. To prove the second step we, first, show that $R = \{\langle s'(y), sub(s,y)\rangle \mid y \in X'\}$ is a bisimulation on sets. We only show the condition concerning urelements, the other conditions can easily be checked.
Assume $\langle s'(x), sub(s,x)\rangle \in R$. We have to prove that it holds: $s'(x) \cap \mathcal{U} = sub(s,x) \cap \mathcal{U}$. Let y be an element of $s'(x) \cap \mathcal{U}$. Then, there are two possibilities. First, $y \in x \cap A$, because s' is a solution of a flat system, and $y = sub(s,y)$ ($y \notin dom(s)$ according to our definition). Second, $y \in e(x) \cap \mathcal{U}$ and additionally $y = sub(s,y)$ (because $y \notin X'$).
Concerning the other direction, we assume y to be an element of $sub(s,x) \cap \mathcal{U}$. Since $y \in X'$, $sub(s,x)$ takes sets as values, therefore $sub(s,x') \in \mathcal{U}$ implies

$x \notin dom(s)$ for $x' \in X$. Conclude: $y = sub(s, y)$, or in other words: $y \notin dom(s)$, and therefore: $y = s'(x)$.
Since R is a bisimulation, we can prove that $s = s' \upharpoonright X$ is a solution for $\mathcal{E}$ by the following equalities. Assume $x \in X$, then the following equalities hold:

$$\begin{aligned}
s(x) &= s'(x) && \text{(Definition of } s(x))\\
&= \{s'(y) \mid y \in e'(x) \cap (X' - X)\} \cup \{s'(y) \mid y \in e'(x) \cap X\}\\
&\quad \cup\ (e' \cap A) && \text{(Definition of a solution s of } \mathcal{E}')\\
&= \{s'(y) \mid y \in e(x) \cap (X' - X)\} \cup \{s'(y) \mid y \in e(x) \cap X\}\\
&\quad \cup\ (e(x) \cap A) && \text{(because } e'(x) = e(x))\\
&= \{sub(s, y) \mid y \in e(x) \cap (X' - X)\} \cup \{sub(s, y) \mid y \in e(x) \cap X\}\\
&\quad \cup\ (e(x) \cap A) && \text{(because } sub(s, y) = s'(y))\\
&= \{sub(s, y) \mid y \in e(x) \cap (X' - X)\} \cup \{sub(s, y) \mid y \in e(x) \cap (X \cup A)\}\\
& && \text{(Logic)}\\
&= \{sub(s, e(x))\}
\end{aligned}$$

Therefore: if $\mathcal{E} = \langle X, A, e\rangle$ is a general system of equations, then there exists a flat system $\mathcal{E}' = \langle X', A, e'\rangle$, such that it holds: if s' is a solution of $\mathcal{E}'$, then $s := s \upharpoonright X$ is a solution for $\mathcal{E}$. This proves the existence of a solution for $\mathcal{E}$. From the above consideration we know: if s is a solution or $\mathcal{E}$, then the solution s' for $\mathcal{E}'$ extends s. Because s' is necessarily unique, s is unique, too. This suffices to show the theorem. q.e.d.

Remark 11.6.4 (i) The theorem shows that the anti-foundation axiom (defined on flat systems of equations) suffices to solve general system of equations. In a certain sense, the anti-foundation axiom is minimal, because solutions of more general forms of systems of equations follow from this axiom.

(ii) Theorem 11.6.3 makes clear that in order to develop the theory, it is not necessary to use urelements (if we want to work in pure set theory). The usage of urelements as indeterminates has no crucial impact to the theory of hypersets: we can simply reformulate the anti-foundation axiom as the claim that every general system of equations has a unique solution. Clearly, one needs to reformulate the definition of the substitution operation $sub(s, b)$ also, because s is defined on urelements. Fortunately, s can be defined on sets as well.

This finishes the basic theory of non-well-founded sets.[19] Now, we are able to solve quite arbitrary systems of equations, but further generalizations are

[19] One of the most important features of the theory of hypersets is the consistency of ZFC_{afa} relative the consistency of ZFC set theory. We do not present a proof for the existence of a model of ZFC_{afa} here. The interested reader is refered to [BarMo96] and [Ac88] for further information concerning this point.

possible. To present these generalizations we leave the ground of elementary mathematics, and more advanced mathematical tools and techniques are required. In particular, we will work in the framework of coalgebras and we will use category theoretic methods in order to develop the theory of coalgebras. The next chapter (dealing with topics of category theory) can be considered as providing the prerequisites for the development of the further theory.

11.7 History

The idea for the introduction of an anti-foundation axiom is relatively old. The work of Finsler dates back to the first half of the 20-th century ([Fi75]). Another account was proposed by Forti and Honsell in [FoHo83]. The theory of non-well-founded sets became popular with the work of Peter Aczel in [Ac88]. In [Ac88], a concise theory was established. Unfortunately, the book is not very simply written and not very explicit: a lot of work remains to be done by the readers themselves. Further examinations were provided by Rutten and Turi in [RuTu93] and others. Our presentation is very close to the development of the theory presented in [BarMo96]. Unfortunately, Theorem 11.6.2 contains a non-trivial flaw in [BarMo96]. Hopefully, we formulated a correct proof of this important theorem on which the complete theory in [BarMo96] as well as aspects in [MoSel96] are based. The concept 'bisimulation' was originally introduced by Johan van Benthem in [Be78]. It was around 1980 when bisimulations began to become more and more important in many different areas in logic and computer science. Later, we will consider further properties of bisimulations. In particular, we will see several further versions of bisimulations adjusted to certain environments.

Chapter 12

Category Theory

This chapter provides necessary background information in category theory. Categories can be used in order to model circularity in a very general framework, the so-called theory of coalgebras. Using category theory as mathematical foundation and the theory of coalgebras as formal framework it will become possible to model certain generalizations of non-well-founded set theory (for example the corecursion theorem), as well as knowledge representations and the distinction between common ground and private knowledge. In developing this, we are forced to deal with quite abstract objects which require quite abstract theories. We will begin our considerations with some basic concepts in category theory like products, coproducts, and natural transformations. After that we will discuss some examples in length. These examples include data structures, term algebras and logical applications. Finally, we will examine certain properties of final objects in a given category. In the context of the next chapters (Chapters 13 and 14), it will become clear why the work in category theory is necessary.

12.1 Categories, Functors, and Operators

First, we need to define the concept of a category. It seems to be hard at first sight to see that the following definition has any reasonable applications. As we will see later, category theory is a theory with a huge variety of applications and examples, in particular because of the generality of the theory.

Definition 12.1.1 *We call $\mathcal{A}$ a category, if $\mathcal{A}$ is a six-tuple of the form*

$$\mathcal{A} = \langle Obj_{\mathcal{A}}, Ar_{\mathcal{A}}, \circ_{\mathcal{A}}, dom, cod, id \rangle$$

such that the following conditions hold:

- *$Obj_{\mathcal{A}}$ is a class of objects of $\mathcal{A}$.*

- *$Ar_{\mathcal{A}}$ is a class of arrows (morphisms) between objects of $\mathcal{A}$. It is convenient to write $[a, b]_{\mathcal{A}}$ for the class of all arrows between object a and object b. Clearly, it is allowed that $[a, b]_{\mathcal{A}}$ is a set.*

- *$\circ_{\mathcal{A}}$ is a binary operation (concatenation) defined on arrows of $\mathcal{A}$, such that $\circ_{\mathcal{A}}$ is associative. Hence it holds: if $f \circ g$ and $g \circ h$ are defined in category $\mathcal{A}$, then it holds: $(f \circ g) \circ h = f \circ (g \circ h)$.*[1]

- *For every object $a \in \mathcal{A}$, $id_a \in [a, a]_{\mathcal{A}}$ such that for an arrow $f : b \longrightarrow a$ and an arrow $g : a \longrightarrow b$ it holds: $id_a \circ f = f$ and $g \circ id_a = g$. id_a is called the identity mapping from a into a.*

- *dom and cod are functions defined on arrows with range in $Obj_{\mathcal{A}}$. We call $dom(f)$ the domain of f, and $cod(f)$ the codomain of f.*

Remark 12.1.1 (i) In category theory, it is important to distinguish between classes and sets. In our definition, we allow that the objects and the arrows of a category are proper classes. Standardly, a category is called large, if the collection of all objects of the category is a proper class. In the other case, namely if the collection of all objects is a set, the category is called (globally) small. If the collection of morphisms $[a, b]_{\mathcal{A}}$ is a set for all pairs $\langle a, b\rangle$ where $a, b \in Obj_{\mathcal{A}}$, then we call this category locally small. For our purposes it is most times not important to distinguish between large and small categories.

(ii) Sometimes we will write simply *Obj*, *Ar* etc. without subscripts, if the context makes it clear to which category we are referring to.

(iii) Notice that for an object $a \in Obj_{\mathcal{A}}$ it holds: id_a is unique. Assume id_a and $\mathtt{id}_a$ are both identities of object a. Because of the defining conditions for identities of Definition 12.1.1 we have the following equalities:

$$id_a = id_a \circ \mathtt{id}_a = \mathtt{id}_a$$

Hence, id_a is uniquely specified.

We would like to develop a concept that enables us not only to speak about mappings (morphisms) between objects of a category, but also about mappings between categories. The appropriate concept is called functor. The following definition introduces this notion.

Definition 12.1.2 *Let $\mathcal{A}$ and $\mathcal{B}$ be two given categories. A functor $\Phi : \mathcal{A} \longrightarrow \mathcal{B}$ is a mapping, such that the following conditions hold:*

(i) $\Phi : Obj_{\mathcal{A}} \longrightarrow Obj_{\mathcal{B}}$

(ii) $\Phi : Ar_{\mathcal{A}} \longrightarrow Ar_{\mathcal{B}}$ *satisfies the condition* $\Phi(f \circ_{\mathcal{A}} g) = \Phi(f) \circ_{\mathcal{B}} \Phi(g)$ *and if additionally* $f \in Ar_{\mathcal{A}}$*, then* $\Phi(f) \in Ar_{\mathcal{B}}$*.*

(iii) For every object $a \in Obj_{\mathcal{A}}$ *it holds:* $\Phi(id_a) = id_{\Phi(a)}$

[1] We suppressed the subscript in $\circ_{\mathcal{A}}$, because it is clear in which category we work. We will do this quite often, when the reference to a particular category is clear.

Remark 12.1.2 (i) A functor Φ can be understood as a structure preserving mapping from a category $\mathcal{A}$ into a category $\mathcal{B}$. The structure preserving nature is essentially captured in the homomorphism-like conditions (ii) and (iii) of Definition 12.1.2.

(ii) Notice that functors can be understood as a pair of mappings $\langle \phi, \psi \rangle$, such that $\phi : Obj_{\mathcal{A}} \longrightarrow Obj_{\mathcal{B}}$ and $\psi : Ar_{\mathcal{A}} \longrightarrow Ar_{\mathcal{B}}$. Notice further that both mappings are connected via the structure preserving properties of functors. ϕ and ψ are not independent of each other.

(iii) In principal, categories can be described without specifying objects at all. This is possible because objects are determined by the identity morphisms id_a. In this respect, objects of a category can be represented as arrows and category theory can be reduced to a theory of arrows.

We will give certain examples in order to clarify the introduced concepts. The following examples are taken from different disciplines of mathematics and provide a first idea of the generality of category theory.

Example 12.1.3 (i) One of the easiest and illuminating examples for categories are partially ordered sets. Assume $\mathbf{D} = \langle D, \leq \rangle$ is a given partially ordered set. We can interpret $\mathbf{D}$ as a category by the associating the following objects. Interpret D as $Obj_{\mathbf{D}}$, and the order relation $\leq$ as arrows $Ar_{\mathcal{D}}$: for all $a, b \in Obj_{\mathcal{D}}$ the expression $a \leq b$ denotes a morphism $f \in Ar_{\mathbf{D}}$, such that $f : a \longrightarrow b$. Every identity arrow id_a for objects in $\mathbf{D}$ exists, because $\leq$ is reflexive. Associativity of concatenation of arrows holds because of the transitivity of $\leq$. Notice that $\mathbf{D}$ is a small category.

(ii) The objects of the category $\mathcal{SET}$ are sets of a given universe (for example ZFC_{afa} or ZFC), the arrows are arbitrary functions between two sets, and concatenation of arrows is ordinary concatenation of functions on sets. The identity mapping on a set a can be interpreted as the identity arrow of an object $a \in Obj_{\mathcal{SET}}$. Notice that $\mathcal{SET}$ is a large category, i.e. $\mathcal{SET}$ is a proper class. On the other hand $\mathcal{SET}$ is locally small, because the collection of all functions between two sets $[a, b]_{\mathcal{SET}}$ is again a set.

(iii) Consider the category of all lattices $\mathcal{LAT}$. Objects are lattices, and arrows are lattice homomorphisms (i.e. order preserving mappings) from a lattice a into a lattice b. The other conditions on $\mathcal{LAT}$ are clear. The category of all groups $\mathcal{GROUP}$ is defined similarly. Here, objects are groups, and arrows are group homomorphisms. Another example is the category $\mathcal{TOP}$ of all topological spaces, where the arrows are continuous functions from a topological space a into a topological space b. All these categories are large but locally small.

(iv) Naturally the above considerations lead us to the category of all categories (usually denoted by $\mathcal{CAT}$). Here the objects are categories and the arrows are functors between categories. Concatenation is defined as concatenation of functors. Identities are the identity functors of the categories. Clearly, these functors exist, because identities on categories satisfy the conditions required for functors. It is clear that such a category is large. Furthermore, $\mathcal{CAT}$ is not locally small in general.

(v) Let us consider some examples of functors. Assume two partially ordered sets $\mathbf{D} = \langle D, \leq \rangle$ and $\mathbf{D}' = \langle D', \leq \rangle$ are considered as categories. Then any order preserving mapping $\Phi : D \longrightarrow D'$ is a functor. If $\mathbf{D}$ and $\mathbf{D}'$ are lattices any lattice homomorphism mapping D into D' is a functor.

(vi) An important functor in category theory is the so-called forgetful functor. Given two categories $\mathcal{A}$ and $\mathcal{B}$ where $\mathcal{B}$ has less structure than $\mathcal{A}$, then every functor $\Phi : \mathcal{A} \longrightarrow \mathcal{B}$ is called a forgetful functor. An easy example is the functor Φ mapping the category of all groups $\mathcal{GROUP}$ into $\mathcal{SET}$: Every group in $Obj_{\mathcal{GROUP}}$ is mapped to its underlying set and any homomorphism in $Ar_{\mathcal{GROUP}}$ is mapped to its underlying set theoretical function. Clearly the image of every object in $Obj_{\mathcal{GROUP}}$ looses structure. The same is true for every morphism in $\mathcal{GROUP}$. The general idea of a forgetful functor is easily generalized to many algebraic structures.

In mathematics, it is often the case that certain duality principles hold. For example, if a lattice $\mathbf{D} = \langle D, \leq \rangle$ is given, we can define the dual of $\mathbf{D}$ by 'turning the lattice upside down'. To give an example: $\mathbf{D}' = \langle D, \geq \rangle$ is the dual of $\mathbf{D}$. In category theoretic terms, one can represent this idea easily. Transform a given category to its dual by reversing the direction of all arrows. The next definition makes this intuition precise.

Definition 12.1.3 *Assume $\mathcal{A} = \langle Obj_{\mathcal{A}}, Ar_{\mathcal{A}}, \circ_{\mathcal{A}}, dom, cod, id \rangle$ is a given category. The dual category $\mathcal{A}^{\mathcal{OP}}$ of $\mathcal{A}$ is the category with the following properties:*

(i) The objects of $\mathcal{A}^{\mathcal{OP}}$ are the objects of $\mathcal{A}$.

(ii) The arrows of $\mathcal{A}^{\mathcal{OP}}$ are the inverted arrows of $\mathcal{A}$. Therefore, if $f \in Ar_{\mathcal{A}}$, with $f : a \longrightarrow b$, then there is a morphism f^{OP} in the category $\mathcal{A}^{\mathcal{OP}}$ with $f : b \longrightarrow a$.

(iii) Concatenation of arrows in $\mathcal{A}^{\mathcal{OP}}$ is defined as follows: $(f \circ_{\mathcal{A}} g)^{\mathcal{OP}} = g^{\mathcal{OP}} \circ_{\mathcal{A}^{\mathcal{OP}}} f^{\mathcal{OP}}$.

(iv) The identity arrow is defined by $(id_a)^{\mathcal{OP}} = id_a$.

Remark 12.1.4 (i) It is easy to check that the definition of concatenation of arrows in the dual category is associative. From this it follows easily that $\mathcal{A}^{\mathcal{OP}}$ is in fact a category.

(ii) Notice that for a given category $\mathcal{A}$ it holds: $(\mathcal{A}^{\mathcal{OP}})^{\mathcal{OP}} = \mathcal{A}$. This is obvious by Definition 12.1.3.

(iii) We should mention that every category can be mapped via a functor to its dual category. This functor maps objects to themselves and inverts the arrows of the given category.

(iv) It is important to notice that the dual of a category $\mathcal{A}$ in which every morphism is a function does not imply that in the dual category $\mathcal{A}^{\mathcal{OP}}$ every morphism is also a function. It can be the case that these morphisms become relations in the dual category. An example would be the category $\mathcal{SET}$. All arrows in $\mathcal{SET}$ are functions mapping sets into sets. The dual category $\mathcal{SET}^{\mathcal{OP}}$ is a category in which some arrows are relations. For example, consider an arrow $f \in Ar_{\mathcal{SET}}$, such that $f : \{a_1, a_2\} \longrightarrow \{b\} : a_i \longmapsto b$ (for $i \in \{1,2\}$). Then, the dual arrow f^{OP} is the relation $f^{OP} = \{\langle b, a_1\rangle, \langle b, a_2\rangle\}$. Clearly, $f^{\mathcal{OP}}$ is not a function.

We add two more examples for dual categories in order to make the reader more familiar with this concept.[2]

Example 12.1.5 (i) Assume $\mathbf{D} = \langle D, \leq\rangle$ is a partially ordered set. Then, according to Example 12.1.3(i) we can consider $\mathbf{D}$ as a category. The category $\mathbf{D}^{\mathcal{OP}}$ has the same objects as $\mathbf{D}$, but the arrows are inverted. An arrow in $Ar_{\mathbf{D}}$ is an edge in the partially ordered set $\mathbf{D}$. Because $\leq$ is transitive, concatenation is well-defined and associative. For every arrow $f : a \longrightarrow b$ in $Ar_{\mathbf{D}}$ there is an arrow $f^{\mathcal{OP}} : b \longrightarrow a$. The identity arrow $id^{\mathcal{OP}}$ is equal to the corresponding identity arrow in $\mathcal{D}$. Therefore, it holds: $id^{\mathcal{OP}} = id$.

(ii) There are categories $\mathcal{A}$ that are isomorphic to their dual $\mathcal{A}^{\mathcal{OP}}$ (which means the two categories cannot be distinguished). The easiest example for this phenomenon is a category $\mathcal{A}$ that has only identity arrows as arrows. Then, it is obvious that the dual category $\mathcal{A}^{\mathcal{OP}}$ is isomorphic to the original category $\mathcal{A}$. A non-trivial example for a category with this property is an Abelian group interpreted as a category. The reason for the fact that an Abelian group is isomorphic to its dual is the algebraic behavior of an Abelian group: Abelian groups have only one (trivial) automorphism, namely the identity.

One central insight of mathematical considerations is the idea of the importance of principles that allow us to build new objects from given objects. For example, one is used to the fact that in algebra product constructions, (direct) sums, exponentiation operations etc. are very important for the underlying theory. Category theory is not an exception. For example, the natural idea of a product of two categories yields the concept of a bifunctor, i.e. a functor that

[2] Clearly, duality principles play an important role in all parts of mathematics and it is not surprising that there is a variety of possible examples of this concept.

is defined on a product of two categories. Such a functor can easily be modeled by a pair of ordinary functors. We will not consider such natural constructions now, but postpone some of them to later sections. In the next section, we will consider constructions like products defined on objects of a given category.

12.2 Constructions

In this section, we will introduce several important concepts of category theory. We will see that these concepts are natural generalizations of ideas well-known in classical mathematical theories. The first concept we will examine is the concept of a product of two objects in a given category. This is a generalization of the more specific concept of a classical product operation as can be found in set theory, group theory, lattice theory, or ring theory. Similarly to algebraic theories we are able to generalize the concept of a product of two objects to the product of arbitrarily many objects. The corresponding generalization is quite straightforward.

Definition 12.2.1 *Assume $\mathcal{A}$ is a given category. The triple $\langle b \times c, \pi_l, \pi_r \rangle$ where $b, c \in Obj_{\mathcal{A}}$, and $\pi_l : b \times c \longrightarrow b$ and $\pi_r : b \times c \longrightarrow c$ are arrows of $Ar_{\mathcal{A}}$, is called a product if and only if the following condition holds: if there are arrows $f, g \in Ar_{\mathcal{A}}$, such that $f : a \longrightarrow b$ and $g : a \longrightarrow c$, then there is a unique arrow $\langle f, g \rangle : a \longrightarrow b \times c$, such that the following diagram commutes:*

$$\begin{tikzcd}
 & a \arrow[dl, "f"'] \arrow[d, "\langle f, g \rangle"] \arrow[dr, "g"] & \\
b & b \times c \arrow[l, "\pi_l"] \arrow[r, "\pi_r"'] & c
\end{tikzcd}$$

The morphisms π_l and π_r are called projection functions. They are similarly interpreted to ordinary projection functions in algebra theories or recursion theory. The uniqueness condition in Definition 12.2.1 is crucial. Without the uniqueness of $\langle f, g \rangle$, the definition would not give us a unique product of objects: the uniqueness of $\langle f, g \rangle$ implies the uniqueness of the products of two objects a and b (up to isomorphisms). More information concerning this point can be found in Theorem 12.2.2).

Although the definition of a product seems quite unfamiliar we can interpret any product in classical algebraic theories as a category theoretic product construction. Consider the following examples.

Example 12.2.1 (i) Assume the category $\mathcal{SET}$ is given. The product operation in set theory is the ordinary Cartesian product of two sets. Assume b and c

are two sets, then $b \times c = \{\langle x, y\rangle \mid x \in b \wedge y \in c\}$ is the Cartesian product, and π_l and π_r are set theoretic projection functions. It is clear that for given functions $f : a \longrightarrow b$ and $g : a \longrightarrow c$ and for all $a' \in a$ the following two equalities hold:

$$\begin{aligned} f(a') &= \pi_l(\langle f, g\rangle(a')) \\ g(a') &= \pi_r(\langle f, g\rangle(a')) \end{aligned}$$

Hence, the corresponding diagram of Definition 12.2.1 commutes, and the ordinary Cartesian product in set theory is really a product in the category theoretical sense.

(ii) Assume $\mathbf{D} = \langle D, \leq\rangle$ is a lattice. The infimum of two elements $a, b \in D$ is a product operation. Because of the fact that $\inf\{b, c\} \leq b$ and $\inf\{b, c\} \leq c$, the projection functions exist for every object of the lattice. And if $a \leq b$ and $a \leq c$, then $a \leq \inf\{b, c\}$ by the definition of infimum. Hence, there is an arrow $a \leq \{b, c\}$, such that the corresponding diagram of Definition 12.2.1 commutes.

(iii) Assume we want to represent logical conjunction of first-order logic using category theoretic terms. Assume further that the 'existence of a proof' is represented by an arrow $\longrightarrow$ and propositions are modeled by objects. For example, if we can prove that $a \vdash b$, then this can be represented by the arrow $f : a \longrightarrow b$. Assume such a category is given. We represent a logical conjunction in a Gentzen-type calculus where we restrict our rules used in deductions, such that at most one formula can occur on the left side of $\vdash$ and at most one formula can occur on the right side of $\vdash$. Then, conjunction should satisfy the following (well-known) conditions.

$$\frac{a \vdash b \qquad a \vdash c}{a \vdash b \wedge c} \qquad\qquad \frac{a \vdash b \wedge c}{a \vdash b} \qquad\qquad \frac{a \vdash b \wedge c}{a \vdash c}$$

The above conditions can be represented in category theoretic terms as follows. Assume that $\times$ is our symbol for products of objects (as defined in Definition 12.2.1) and assume further that propositions are modeled by objects of a category and provability as an arrow of the same category. Then, the following represents the conditions governing the properties of a conjunction.

$$\frac{f : a \longrightarrow b \qquad g : a \longrightarrow c}{\langle f, g\rangle : a \longrightarrow b \times c} \qquad\qquad \frac{\langle f, g\rangle : a \longrightarrow b \times c}{f : a \longrightarrow b}$$

$$\frac{\langle f, g\rangle : a \longrightarrow b \times c}{f : a \longrightarrow c}$$

The possibility to model proof-theoretic rules in category theoretic terms was the origin of a strong endeavor of using category theory in proof theoretical applications. We cannot follow this development here and refer the reader to

[Tr92, GiLaTa89, Da92] for an overview of the relevance of category theoretic in proof theory.

(iv) In the category of all groups $\mathcal{GROUP}$, the category theoretic product corresponds to the ordinary direct product in group theory. For topological spaces there is the correspondence to topological products. It is possible to transfer product operations of other disciplines of mathematics (such as graph theory, the theory of vector spaces etc.) to category theoretic concepts as well. This exemplifies the generality of Definition 12.2.1.

The next theorem shows that products are unique up to isomorphisms (similar to the cases in our algebraic examples).

Theorem 12.2.2 *Assume a category $\mathcal{A}$ is given. If $\mathcal{A}$ has products, then any two products of two objects $a, b \in Obj_{\mathcal{A}}$ are isomorphic.*

Proof: Assume $a \times b$ and $a \cdot b$ are products with projection functions π_l, π_r, and ϕ_l, ϕ_r, respectively. We have to show that both products are isomorphic. Because $a \times b$ is a product and $\phi_l : a \cdot b \longrightarrow a$ and $\phi_r : a \cdot b \longrightarrow b$, it holds: $\langle \phi_l, \phi_r \rangle : a \cdot b \longrightarrow a \times b$. Moreover, the morphism $\langle \phi_l, \phi_r \rangle$ is unique. Similarly, it holds that $\langle \pi_l, \pi_r \rangle : a \times b \longrightarrow a \cdot b$ is unique. Then, it follows that $\langle \pi_l, \pi_r \rangle \circ \langle \phi_l, \phi_r \rangle$ and $\langle \phi_l, \phi_r \rangle \circ \langle \pi_l, \pi_r \rangle$ is the identity arrow. This implies that both products are isomorphic. q.e.d.

A natural idea is to generalize products of two objects to products of arbitrarily many objects. This is quite similar to the generalization of the Cartesian product of two sets $a \times b$ to the product $\prod_{i \in I} a_i$ where I is a given index set and all a_is are sets. We state the general definition of products in order to be complete.

Definition 12.2.3 *Assume that $\mathcal{A}$ is a given category. Assume further that there is a family $\{b_i\}_{i \in I}$ of objects of $\mathcal{A}$ where I is an index set. The pair $\langle \prod_{i \in I} b_i, \{\pi_i\}_{i \in I} \rangle$, such that for every $j \in I$ it holds: $b_j \in Obj_{\mathcal{A}}$ and $\pi_j : \prod_{i \in I} b_i \longrightarrow b_j$ is called a product if and only if the following condition holds: if there are arrows $f_k : a \longrightarrow b_k$ for every $k \in I$, then there is a unique arrow $f : a \longrightarrow \prod_{i \in I} b_i$ such that $\pi_k f = f_k$.*

It is quite easy to draw a picture for the finite case to get a better idea how the general case Definition 12.2.3 works. Notice that Definition 12.2.3 is a natural generalization of Definition 12.2.1. For the infinite case we cannot draw pictures, but the above definition does work for the infinite case as well.

Although it seems to be the case that every category allows the introduction of products this is not true. For example, partially ordered sets $\mathbf{D} = \langle D, \leq \rangle$ interpreted as categories, do not have products in general, because the infimum of arbitrary elements $x, y \in D$ is not generally defined.

In Example 12.1.3, we saw that there is the possibility to define the dual of a given category $\mathcal{A}$. This construction is determined by inverting the arrows in $\mathcal{A}$. Duality principles are quite common in mathematics. The same is true for category theory. Here, every dual construction and every dual object of a given category is denoted by co- followed by the name of the construction or the object. For example, consider products. The dual construction is called a coproduct. The next definition makes the concept of a coproduct precise.

Definition 12.2.4 *Assume a category $\mathcal{A}$ is given. A triple $\langle b{+}c, in_l, in_r\rangle$ where $b{+}c \in Obj_{\mathcal{A}}$ and $in_l : b \longrightarrow b{+}c$ and $in_r : c \longrightarrow b{+}c$ are arrows in $\mathcal{A}$, is called a coproduct (or sum) if and only if the following condition holds: if there are arrows $f, g \in Ar_{\mathcal{A}}$, such that $f : b \longrightarrow a$ and $g : c \longrightarrow a$, then there is a unique arrow $\langle f, g\rangle : b + c \longrightarrow a$, such that the following diagram commutes:*

$$\begin{tikzcd}
 & a & \\
b \arrow[ur, "f"] \arrow[r, "in_l"'] & b+c \arrow[u, "\langle f,g\rangle"'] & c \arrow[ul, "g"'] \arrow[l, "in_r"]
\end{tikzcd}$$

Remark 12.2.2 (i) A coproduct can be naturally associated with the dual construction of a product. (Compare the diagrams in Definition 12.2.1 and Definition 12.2.4.) The inverted morphisms of the projection functions in the above definition of a coproduct are usually called injections.

(ii) Alternative names for coproducts are sums and directed sums (dependent on the algebraic theory). This can be motivated by the fact that coproducts are generalizations of sums used in many algebraic theories (compare Example 12.2.3). Therefore, our usage of the symbol '+' in order to represent a coproduct is an intuitive usage for the construction but does not represent the duality principle behind the construction.

We will add some examples for coproducts in the following. Again these constructions are familiar from classical algebraic theories.

Example 12.2.3 (i) Consider again the category $\mathcal{SET}$. Whereas, in the case of products the corresponding set theoretic construction was the Cartesian product, the associated construction in set theory for coproducts is disjoint union. Assume $b, c \in Obj_{\mathcal{SET}}$. Then the triple

$$\langle\{\langle x, i\rangle \mid (x \in b \wedge i = 0) \vee (x \in c \wedge i = 1)\}, in_l, in_r\rangle$$

where $in_l : b \longrightarrow \{\langle b', 0\rangle \mid b' \in b\}$ and $in_r : c \longrightarrow \{\langle c', 1\rangle \mid c' \in c\}$ is a coproduct. It is easily checked that the diagram in Definition 12.2.4 commutes for this set theoretical example. The disjoint union in set theory is the prototypical example for a coproduct.

(ii) Assume $\mathbf{D} = \langle D, \leq\rangle$ is a lattice interpreted as a category. We saw in Example 12.2.1(ii) that products in lattices correspond to infima of objects. Similarly, a coproduct of a lattice is given by the supremum of objects of that category. This can be checked immediately.

(iii) For the standard algebraic structures, coproducts are constructions that are associated with one or the other form of sums. In the category of topological spaces, topological sums satisfy the coproduct conditions and in the category of groups we can associate free sums with coproducts. In the category of Abelian groups as well as in the category of R-modules, direct sums are examples of coproducts.

(iv) In Example 12.2.1(iii), we saw how to model conjunctions and their behavior in category theoretic terms. We want to do the same for coproducts. The corresponding logical connective for coproducts is disjunction (interpreted as an exclusive disjunction). The corresponding rules for disjunction can be specified as follows.

$$\frac{a \vdash c \qquad b \vdash c}{a \vee b \vdash c} \qquad\qquad \frac{a \vdash b}{a \vdash b \vee c} \qquad\qquad \frac{a \vdash c}{a \vdash b \vee c}$$

The translation of these rules into category theoretic terms is quite straightforward. Again, proofs are interpreted as arrows and propositions as objects. Then we get the following corresponding rules.

$$\frac{f : a \longrightarrow c \qquad g : b \longrightarrow c}{\langle f, g\rangle : a + b \longrightarrow c} \qquad\qquad \frac{f : a \longrightarrow b}{in_l \circ f : a \longrightarrow b + c}$$

$$\frac{f : a \longrightarrow c}{in_r \circ f : a \longrightarrow b + c}$$

Similarly to the case of products one can show that coproducts are unique up to isomorphisms. The next theorem states this fact.

Theorem 12.2.5 *Assume a category $\mathcal{A}$ is given, furthermore $\mathcal{A}$ allows coproducts. Then it holds: any two coproducts of objects $a, b \in Obj_{\mathcal{A}}$ are isomorphic.*

Proof: Assume $a+b$ and $a \oplus b$ are coproducts with injection arrows in_l, in_r, and ϕ_l, ϕ_r, respectively. Because $a + b$ is a coproduct, $\phi_l : a \longrightarrow a \oplus b$ and $\phi_r : b \longrightarrow a \oplus b$, we have (similar to the case in Theorem 12.2.2) the unique morphism $\langle \phi_l, \phi_r\rangle : a + b \longrightarrow a \oplus b$. Uniqueness holds also for the morphism

$\langle in_l, in_r \rangle : a \oplus b \longrightarrow a + b$. Therefore, $\langle \phi_l, \phi_r \rangle \circ \langle in_l, in_r \rangle$ as well as the dual arrow $\langle in_l, in_r \rangle \circ \langle \phi_l, \phi_r \rangle$ are the identity arrow. Conclude: every two coproducts are isomorphic. q.e.d.

The natural generalization of coproducts of two objects to coproducts of arbitrarily many objects of a given category can be achieved in a similar way as it was described in Definition 12.2.3 for the generalization of products. The precise definition of the general form of the definition of coproducts is formulated in the next definition.

Definition 12.2.6 *Assume that $\mathcal{A}$ is a given category. Assume further that there is a family $\{b_i\}_{i \in I}$ of objects of $\mathcal{A}$ where I is an index set. The pair $\langle \coprod_{i \in I} b_i, \{in_i\}_{i \in I} \rangle$, such that for every $j \in I$ it holds: $b_j \in Obj_{\mathcal{A}}$ and $in_j : b_j \longrightarrow \coprod_{i \in I} b_i$, is called a coproduct if and only if the following condition holds: if for every $j \in I$ there are arrows $f_j : b_j \longrightarrow a$, then there is a unique arrow $f : \coprod_{i \in I} b_i \longrightarrow a$ such that $f \circ in_j = f_j$.*

The next category theoretic concept we introduce is the concept of a pullback. Pullbacks will become important in the theory of coalgebras, in particular when we will model infinite streams.

Definition 12.2.7 *Assume $\mathcal{A}$ is a given category. Assume further that two arrows $f : a \longrightarrow c$ and $g : b \longrightarrow c$ are given. A pullback (sometimes also called Cartesian square) is a triple $\langle d, h, r \rangle$ where $d \in Obj_{\mathcal{A}}$, and $h : d \longrightarrow a$ and $r : d \longrightarrow b$ are arrows of $\mathcal{A}$, such that the following condition holds. If $u : d' \longrightarrow a$ and $v : d' \longrightarrow b$ are two arrows of $Ar_{\mathcal{A}}$, such that $f \circ u = g \circ v$, then there exists exactly one arrow $g : d' \longrightarrow d$, such that the following diagram commutes:*

$$\begin{tikzcd}
d' \arrow[dr, "\gamma"] \arrow[drr, bend left, "v"] \arrow[ddr, bend right, "u"'] & & \\
 & d \arrow[r, "r"] \arrow[d, "h"'] & b \arrow[d, "g"] \\
 & a \arrow[r, "f"'] & c
\end{tikzcd}$$

The basic idea of pullbacks can be summarized as follows: we can define an arrow u with range a and an arrow v with range b that are induced by an arrow f with range c and an arrow g with range c. An important property is that if f is a monomorphism, then the newly defined arrow u is a monomorphism, too. We want prove this property of pullbacks now.

Theorem 12.2.8 *Assume that a category $\mathcal{A}$ is given. Assume further that $\langle d, h, r\rangle$ is a pullback in $\mathcal{A}$ as depicted in the diagram of Definition 12.2.7. Then it holds: if f is a monomorphism, then r is a monomorphism, too.*

Proof: Assume that $\langle d, h, r\rangle$ is a pullback as in the diagram of Definition 12.2.7. Assume further that $f : a \longrightarrow c$ is a monomorphism, and that γ_1 and γ_2 are arrows with $\gamma_1 : d' \longrightarrow d$ and $\gamma_2 : d' \longrightarrow d$, such that $r \circ \gamma_1 = r \circ \gamma_2$. It follows that $g \circ r \circ \gamma_1 = g \circ r \circ \gamma_2$, and because $g \circ r = f \circ h$, we get $f \circ h \circ \gamma_1 = f \circ h \circ \gamma_2$. Because f is a monomorphism, we can infer $h \circ \gamma_1 = h \circ \gamma_2$. Hence, $\gamma_1 = \gamma_2$, because $\langle d, h, r\rangle$ is a pullback. This implies that r must also be a monomorphism. q.e.d.

In the following, we will mention some examples of pullbacks. In the theory of coalgebras, we will see more examples for this concept.

Example 12.2.4 (i) Consider the category $\mathcal{SET}$. An example for a pullback in $\mathcal{SET}$ is the triple $\langle a \otimes b, f, g\rangle$, such that $a \otimes b$ is defined as follows (where $a \times b$ represents the ordinary Cartesian product):

$$a \otimes b = \{\langle x, y\rangle \in a \times b \mid f(x) = g(y)\}$$

Assume additionally that $h : a \otimes b \longrightarrow a$ and $g : a \otimes b \longrightarrow b$ are ordinary projection functions (because $a \otimes b$ consists of pairs of elements of the Cartesian product of a and b projection functions are well-defined in this context). To see that the triple $\langle a \otimes b, f, g\rangle$ is in fact a pullback, consider the following. First, notice that the following diagram commutes.

$$\begin{tikzcd}
a \otimes b \arrow[r, "r"] \arrow[d, "h"'] & b \arrow[d, "g"] \\
a \arrow[r, "f"'] & c
\end{tikzcd}$$

If we map an arbitrary pair $\langle x, y\rangle \in a \otimes b$ via h to a, then $h(\langle x, y\rangle) = x$. For the application of r it holds: $r(\langle x, y\rangle) = y$. Now, our condition $f(x) = g(y)$ on $a \otimes b$ ensures that $f \circ h = g \circ r$. To see that for an arbitrary triple $\langle d', u, v\rangle$ which satisfies the corresponding conditions of $\langle a \otimes b, h, r\rangle$ (i.e. which yields a commutative diagram when $\langle a \otimes b, h, r\rangle$ is replaced by $\langle d', u, v\rangle$) there is a unique mapping $\gamma : d' \longrightarrow a \otimes b$, such that the corresponding diagram commutes, consider the following. Choose an arbitrary element $e \in d'$. Then, $u(e) = x$ and $v(e) = y$. Because the diagram for $\langle d', u, v\rangle$ commutes, we have $g(y) = z$ and $f(x) = z$. Hence, we have a uniquely defined pair $\langle x, y\rangle \in a \otimes b$, such that $\gamma(e) = \langle x, y\rangle$. This shows that γ is unique.

(ii) Consider the category of all groups. An example for a pullback in the category of all groups is quite similar to our first example (i). Given two arrows $f : a \longrightarrow c$ and $g : b \longrightarrow c$ (a, b, c are groups and f, g are group homomorphisms) we can construct a pullback as a triple $\langle d, h, r \rangle$, such that d is a subgroup of the product of the groups a and b, such that it holds: $\langle x, y \rangle \in a \times b$ if and only if $f(x) = g(y)$. It is easy to check that this specification determines a pullback. Notice the similarity of the construction to (i).

(iii) In topological spaces, there is already a well-known construction that corresponds to pullbacks. Fiber produces of given topological spaces a and b are precisely pullbacks where the induced topology is given by the product topology. The connection to topology is the reason that sometimes pullbacks are called fiber products.

We introduce a slightly different type of pullback, a so-called weak pullback. Although weak pullbacks differ only in the uniqueness condition of the existence of the morphism γ (compare Definition 12.2.7), we state the definition fully explicit here.

Definition 12.2.9 *Assume $\mathcal{A}$ is a given category. Assume further that two arrows $f : a \longrightarrow c$ and $g : b \longrightarrow c$ are given. A weak pullback is a triple $\langle d, h, r \rangle$ where $d \in Obj_{\mathcal{A}}$, and $h : d \longrightarrow a$ and $r : d \longrightarrow b$ are arrows of $\mathcal{A}$, such that the following condition holds. If $u : d' \longrightarrow a$ and $v : d' \longrightarrow b$ are two arrows of $\mathcal{A}$, such that $f \circ u = g \circ v$, then there exists at least one arrow $\gamma : d' \longrightarrow d$, such that the following diagram commutes:*

$$\begin{tikzcd}
d' \arrow[dr, "\gamma"] \arrow[drr, bend left, "v"] \arrow[ddr, bend right, "u"'] & & \\
& d \arrow[r, "r"] \arrow[d, "h"'] & b \arrow[d, "g"] \\
& a \arrow[r, "f"'] & c
\end{tikzcd}$$

Similar to the case of products, there is also for pullbacks a dual construction, a so-called pushout. This co-construction consists essentially in inverting the arrows. It is the generality of category theory that enables us to speak about the dual construction of a pullback as a concept that just inverts the arrows of a product. Nevertheless, we want to define pushouts explicitly here.

Definition 12.2.10 *Assume a category $\mathcal{A}$ is given. Assume further that two arrows $f : a \longrightarrow b$ and $g : a \longrightarrow c$ are given. A pushout (sometimes called* *a cocartesian square) is a triple $\langle d, h, r\rangle$, where $d \in Obj_{\mathcal{A}}$, $h : b \longrightarrow d$ and $r : c \longrightarrow d$ are arrows of $Ar_{\mathcal{A}}$, and it holds $h \circ f = r \circ g$, such that the following condition holds: If there is an object d' and arrows $u : b \longrightarrow d'$ and $v : c \longrightarrow d'$, such that $u \circ f = v \circ g$, then there is a unique arrow $\gamma : d \longrightarrow d'$, such that $\gamma \circ h = u$ and $\gamma \circ g = v$. In other words: the following diagram commutes:*

$$\begin{tikzcd}
a \arrow[r, "f"] \arrow[d, "g"'] & b \arrow[d, "h"] \arrow[rdd, bend left, "u"] & \\
c \arrow[r, "r"'] \arrow[rrd, bend right, "v"'] & d \arrow[rd, "\gamma"] & \\
& & d'
\end{tikzcd}$$

In order to give a more concrete idea of the properties of pushouts, we add some examples for this concept.

Example 12.2.5 (i) Pushouts are dual constructions of pullbacks. Consider the category $\mathcal{SET}$ again. Assume that two sets b and c together with arrows $h : b \longrightarrow d$ and $r : c \longrightarrow d$ are given. Assume further that a is equal to $b \cap c$ and that f and g are inclusion arrows. Then, the triple $\langle b \cap c, id_b, id_c\rangle$ is a pushout as can be easily verified.

(ii) Assume we have the same situation as in (i), except that a is not equal to the intersection of b and c. First, take the disjoint union $b+c$ of sets b and c and define an equivalence relation $\sim$ on $b + c$, according to the following condition:

$$\forall x \in b \forall y \in c : (x \sim y \ \leftrightarrow \ \exists z \in a : f(z) = x \ \wedge \ g(z) = y)$$

Then $\langle b + c/_{\sim}, in_r, in_l\rangle$ is a pushout.

(iii) Concerning topological spaces, pushouts correspond to fiber sums of two given topological spaces. The topology on d is induced by the topology on the fiber sum. Like in the case of pullbacks the concept of a pushout originates from topology.

(iv) In the category of all groups, a pushout of two groups b and c is the group $b \cdot c/_{z}$ where z is the smallest normal subgroup in $b \cdot c$. (Here, $b \cdot c$ represents the free product of a and b, i.e. the ordinary product in the category theoretic sense.)

We finish this section about constructions with these concepts. Further remarks should be added concerning a common property of all these constructions. In category theory, products and pullbacks together with their co-counterparts are special forms of limits (colimites). We do not have the space to say more about interesting generalizations of the presented concepts. In the next section, we will introduce two very important notions: natural transformations, a concept that was formulated at the very beginning of category theory and adjunctions, probably the most important concept of category theory, strongly used in logic and computer sciences.

12.3 Natural Transformations and Adjunctions

In the last section, we examined constructions on objects and arrows. Whereas we worked purely in a given category the possibility exists to define constructions on categories and functors as well. In this section, we will introduce two of these operations. Natural transformations are in a certain sense one of the most fundamental ideas in category theory. They played a prominent role when category theory was developed. The next definition introduces this concept precisely.

Definition 12.3.1 *Assume two categories $\mathcal{A}$ and $\mathcal{B}$ are given. Assume further that F and G are two functors with $F : A \longrightarrow B$ and $G : A \longrightarrow B$. A natural transformation $\alpha : F \longrightarrow G$ is a function that maps every object $a \in Obj_{\mathcal{A}}$ to an arrow $\alpha(a) = F(a) \longrightarrow G(a) \in [F(a), G(a)]$, such that the following condition hold: For every $f \in Ar_{\mathcal{A}}$ with $f : a \longrightarrow b$ it holds: $\alpha(b) \circ F(f) = G(f) \circ \alpha(a)$ is a morphism in $Ar_{\mathcal{B}}$ mapping $F(a)$ into $G(b)$.*

Remark 12.3.1 (i) It is easier to use a diagram in order to clarify the situation. The crucial conditions for natural transformations can be represented as a commuting diagram. Notice that all arrows and all objects of the following diagram are in $\mathcal{B}$.

$$\begin{tikzcd}
F(a) \arrow[r, "\alpha(a)"] \arrow[d, "F(f)"'] & G(a) \arrow[d, "G(f)"] \\
F(b) \arrow[r, "\alpha(b)"'] & G(b)
\end{tikzcd}$$

Natural transformations are defined as functions mapping objects of $\mathcal{A}$ into arrows of $\mathcal{B}$. As we can see in the diagram above, $\alpha(a)$ maps for every object $a \in Obj_{\mathcal{A}}$ an object $F(a) \in Obj_{\mathcal{B}}$ into an object $G(a) \in Obj_{\mathcal{B}}$. In other words: α maps F into G, such that the structure preserving properties of F and G

are preserved.

(ii) Natural transformations mark the starting point of category theory. This concept was introduced in [EiMa45] as one of the first concepts at the very beginning of category theory.

We will consider two examples.

Example 12.3.2 (i) Consider $\mathbf{A} = \langle A, \leq \rangle$ and $\mathbf{B} = \langle B, \leq' \rangle$ are partially ordered sets considered as categories. Assume further that $F : \mathbf{A} \longrightarrow \mathbf{B}$ and $G : \mathbf{A} \longrightarrow \mathbf{B}$ are functors, i.e. F and G are order preserving mappings. If $\alpha : F \longrightarrow G$ is a natural transformation, then it holds: $F(a) \leq' G(a)$. This is true, because there is only one arrow from $F(a)$ to $G(a)$ and therefore this arrow is equal to $F(a) \leq' G(a)$. This implies that there is a natural transformation just in case $F \leq' G$, i.e. for every $a \in Obj_{\mathbf{A}}$ it holds $F(a) \leq' G(a)$ (the order is induced pointwise). It is clear that the corresponding diagram of Remark 12.3.1(i) commutes in that case.

(ii) Consider the category of all lattices $\mathcal{LAT}$. Assume that there are two lattices a and b given. The product of a lattice x with a lattice a, can be viewed as a functor mapping $\mathcal{LAT}$ into $\mathcal{LAT}$. Consider $F_a : \mathcal{LAT} \longrightarrow \mathcal{LAT} : x \longmapsto x \times a$ and $G_b : \mathcal{LAT} \longrightarrow \mathcal{LAT} : x \longmapsto x \times b$. What happens with morphisms in $\mathcal{LAT}$, if F_a and G_b are applied? If $f : d \longrightarrow d'$ is a lattice homomorphism in $\mathcal{LAT}$, then the natural choice is captured by the following conditions:

$$F_a(f) : \mathcal{LAT} \longrightarrow \mathcal{LAT} : f \longmapsto \langle f, id_a \rangle$$

$$G_b(f) : \mathcal{LAT} \longrightarrow \mathcal{LAT} : f \longmapsto \langle f, id_b \rangle$$

Notice that the functor F_a maps an arrow $f : d \longrightarrow d'$ into a pair because the lattice d is mapped into $d \times a$ and d' is mapped into $d' \times a$. Therefore, we need two arrows for $F_a(f)$ to make sure that the arrow is defined on $d \times a$. Now, an example for a natural transformation on this category of products of lattices can be easily given by 'inverting the pairs'. Given an arrow $f : d \longrightarrow d'$ a natural transformation $\alpha(f)$ can be defined according to: $\alpha(f)(x) : x \times d \longrightarrow x \times d'$. Then, the following identification is justified: $\alpha(f)(x) = \langle id_x, f \rangle$.

One of the most important concepts in category theory is the concept of an adjunction. The importance of adjunctions is partially based on various applications of this concept. In fact, we can find adjunctions in many different areas of mathematical discourse. The definition of an adjunction is not very illuminating at first sight, and it is not easy to describe this concept intuitively. Adjunctions associate functors that 'preserve sets of arrows $[a, b]$ between two objects'. In fact, the functors map a category $\mathcal{A}$ into a category $\mathcal{B}$ (and vice versa), such that all homomorphism sets are isomorphic. We shall consider the definition of this important concept. Before we formulate this precisely, we specify what we mean by a natural bijection between two objects a and b.

Definition 12.3.2 *(i) Assume $\mathcal{A}$ and $\mathcal{B}$ are two given categories and $F : \mathcal{A} \longrightarrow \mathcal{B}$ and $G : \mathcal{B} \longrightarrow \mathcal{A}$ are two functors. Assume that there is an isomorphism $\phi : [a, G(b)] \longrightarrow [F(a), b]$, i.e. it holds: $[a, G(b)] \cong [F(a), b]$. Given an arrow $f' : a' \longrightarrow a$ in $Ar_{\mathcal{A}}$ we call this isomorphism ϕ natural in a if and only if the following diagram commutes:*

$$\begin{tikzcd}
{[a, G(b)]} \arrow[r, "\phi_{a,b}"] \arrow[d, "f^*"'] & {[F(a), b]} \arrow[d, "F(f^*)"] \\
{[a', G(b)]} \arrow[r, "\phi_{a',b}"'] & {[F(a', b)]}
\end{tikzcd}$$

In the above diagram, f^* is defined according to the condition: If $g \in [a, G(b)]$, then $f^*(g) = g \circ f$.

(ii) Assume an arrow $g : b \longrightarrow b'$ in $Ar_{\mathcal{B}}$ is given. We call the isomorphism $\phi : [a, G(b)] \longrightarrow [F(a), b]$ natural in b if and only if the following diagram commutes:

$$\begin{tikzcd}
{[a, G(b)]} \arrow[r, "\phi_{a,b}"] \arrow[d, "G(g^*)"'] & {[F(a), b]} \arrow[d, "g^*"] \\
{[a, G(b')]} \arrow[r, "\phi_{a,b'}"'] & {[F(a), b']}
\end{tikzcd}$$

(iii) If for $f : a \longrightarrow a'$ and $g : b \longrightarrow b'$ both diagrams above commute, then we call ϕ natural in a and b.

Using Definition 12.3.2 it is quite easy to define the concept of an adjunction.

Definition 12.3.3 *Assume as above that $\mathcal{A}$ and $\mathcal{B}$ are two given categories and $F : \mathcal{A} \longrightarrow \mathcal{B}$ and $G : \mathcal{B} \longrightarrow \mathcal{A}$ are two functors. We call the triple $\langle F, G, \phi \rangle$ an adjunction if and only if $\phi : [x, G(y)] \longrightarrow [F(x), y]$ is a natural bijection in x and y for all $x \in Obj_{\mathcal{A}}$ and $b \in Obj_{\mathcal{B}}$. The functor F is called the left adjoint to G and G is called the right adjoint to F. ϕ is called the adjunction isomorphism for $\langle F, G \rangle$.*

We add some remarks concerning the concept of an adjunction.

Remark 12.3.3 (i) We will not give examples of the above construction in this section. We refer the reader to Section 12.4 for examples and additional remarks concerning properties and the behavior of adjunctions.

(ii) We will see later that seemingly trivial functors like the forgetful functor induces interesting constructions when we consider these functors in the light of adjunctions. This is quite surprising, because the forgetful functor maps objects to ’less interesting objects’ in the sense that the latter ones have less structure.

(iii) Adjunctions and adjoint functors are very common in mathematics. Although Definition 12.3.3 seems to be very specific we will see that in various parts of mathematics and logic adjunctions play an important role. Sanders MacLane pointed this out with the following statement:

> ”The slogan is: ’Adjoint functors arise everywhere’”.[3]

(iv) Definition 12.3.3 seems to be quite complicated. In fact, the idea behind the whole consideration is relatively simple. This idea is if functors are related by an adjunction, then they can be ’interdefined’, i.e. if we have one of the functors F and G, the other one is determined up to isomorphisms.

The following remark gives an alternative definition of an adjunction.

Remark 12.3.4 There are alternative definitions of adjunctions. We shall mention one definition that uses the concept of a natural transformation. Given two categories $\mathcal{A}$ and $\mathcal{B}$, two functors $F : \mathcal{A} \longrightarrow \mathcal{B}$ and $G : \mathcal{B} \longrightarrow \mathcal{A}$, and a natural transformation $\alpha : I_{\mathcal{A}} \longrightarrow G \circ F$, the triple $\langle F, G, \alpha \rangle$ is called an adjunction if for every $b \in Obj_{\mathcal{B}}$ and for every arrow $f : a \longrightarrow G(b)$ with $f \in Ar_{\mathcal{A}}$ there exists exactly one arrow $f^\sharp : F(a) \longrightarrow b$, such that the following diagram commutes:

$$\begin{tikzcd}
I_{\mathcal{A}(a)} \arrow[r, "\alpha_a"] \arrow[dr, "f"'] & G \circ F(a) \arrow[d, "G(f^\sharp)"] \\
 & G(b)
\end{tikzcd}$$

We can state the following fact relating the two different definitions of adjunctions.

[3] Cf. [Ma71].

Fact 12.3.4 *Definition 12.3.3 of an adjunction and the alternative definition of an adjunction in Remark 12.3.4 are equivalent.*

Proof: "$\Rightarrow$" Assume that $\langle F, G, \phi \rangle$ is a pair that is an adjunction according to Definition 12.3.3. Then, the isomorphism ϕ induces an isomorphism $\phi' : [I(a), I(a')] \longrightarrow [G \circ F(a), G \circ F(a')]$. It follows that $\alpha : I_{\mathcal{A}} \longrightarrow G \circ F$ is a natural transformation. Furthermore, it holds (because of the isomorphism ϕ) that for every $f \in [a, G(b)]$ there is exactly one arrow $f^\sharp : G(a) \longrightarrow b$. The commutativity of the diagram above follows from the fact that ϕ is natural.

"$\Leftarrow$" Trivial. q.e.d.

In the following section, we shall give examples and additional facts concerning the concepts we introduced so far. We will stress the ideas of adjoint functors and natural transformations. With a certain intention we will consider a logical example in more detail: intuitionistic implication and conjunction (as other connectives) form a pair of adjunctions. Features like that make it plausible to use category theory even in the area of substructural logic.

12.4 Further Examples and Additional Facts

In this section, we want to summarize examples of adjunctions and natural transformations. Additionally, we will state some facts and properties of these constructions. With respect to the ideas of Section 12.2 some remarks concerning limits are also included. We will not try to be complete in any sense because this chapter should not serve for a concise and exhaustive introduction into category theory. The intention is simply to clarify some concepts that will occur later in the development of the theories of coalgebras, corecursive definitions and circular set theory.

We begin our considerations with some easy examples concerning adjunctions. Examples of adjunctions are probably more important for an understanding of the concept 'adjunction', than the literal definition itself.

12.4.1 A Remark on Partial Orders

Assume two partially ordered sets $\mathbf{D} = \langle D, \leq \rangle$ and $\mathbf{D'} = \langle D', \leq \rangle$ are given. We can interpret $\mathbf{D}$ and $\mathbf{D'}$ as categories where for every $a, b \in D$ it holds: $f : a \longrightarrow b$ if and only if $a \leq b$. A similar relation holds also for $\mathbf{D'}$. A functor $F : \mathbf{D} \longrightarrow \mathbf{D'}$ preserves the order, therefore it must hold:

$$a \leq b \longrightarrow F(a) \leq F(b)$$

This is simply the monotonicity condition for functors. That means, if F preserves the order relation, then F has the property to be monotone in the $\leq$ relation and vice verse.

What is an adjunction? Consider the triple $\langle F, G, \alpha \rangle$. We can state the following condition: $\langle F, G, \alpha \rangle$ is an adjunction if $F : \mathbf{D} \longrightarrow \mathbf{D}'$ and $G : \mathbf{D}' \longrightarrow \mathbf{D}$ are functors, such that the mapping $\alpha : I_{\mathbf{D}}(a) \longrightarrow G \circ F(a)$ is a natural transformation.[4] This is true, if for every $a \in Obj_{\mathbf{D}}$ it holds: $a \leq G(F(a))$. This is our first requirement for the triple $\langle F, G, \alpha \rangle$, in order to justify the properties of an adjunction. We need to make sure to satisfy a second condition: It must hold that for every arrow $f : a \longrightarrow G(b) \in Ar_{\mathbf{D}}$ there exists exactly one arrow $f^{\sharp} : F(a) \longrightarrow b$, such that the corresponding diagram in Remark 12.3.4 commutes. This translates to the condition $F(G(b)) \leq b$ for every $b \in Obj_{\mathbf{D}'}$. We state a justification for this claim: We set $a \leq G(b)$ and $f : G(b) \longrightarrow G(b)$ (which is obviously the identity arrow). Then it holds: for every $f : G(a) \longrightarrow G(b)$, there is exactly one $f^{\sharp} : F(G(b)) \longrightarrow b$, provided our relation $F(G(b)) \leq b$ holds. Conclude: The two functors F and G define an adjunction if for every $a \in Obj_{\mathbf{D}}$ it holds: $a \leq G(F(a))$ and for every $b \in Obj_{\mathbf{D}'}$ it holds: $F(G(b)) \leq b$.

12.4.2 Products and Coproducts

We introduced adjunctions as relations between functors, whereas products and coproducts (limits in general) are operations on objects of a given category. Because functors can map a category $\mathcal{A}$ into itself, it is possible to interpret products as adjunctions. Assume $\mathcal{A}$ is a given category. We examine the case of coproducts. We can consider coproducts as a mapping of a functor, such that the following condition holds. Let $F : \mathcal{A} \times \mathcal{A} \longrightarrow \mathcal{A}$ be defined by:

(i) $F(\langle a, b \rangle) = a + b$ for a and b be objects in $\mathcal{A}$

(ii) $F(\langle f : a \longrightarrow a', g : b \longrightarrow b' \rangle) = \langle f, g \rangle : a + b \longrightarrow a' + b'$
where f and g are arrows in $\mathcal{A}$

Our claim can be formulated as follows: F is the left adjoint of the functor $G : \mathcal{A} \longrightarrow \mathcal{A} \times \mathcal{A}$, such that composition and identity is pointwise induced. (The product $\mathcal{A} \times \mathcal{A}$ has as objects pairs of objects of $\mathcal{A}$ and as arrows pairs of arrows of $\mathcal{A}$.) First, it is clear that for every pair of objects $\langle a, b \rangle$, there is an arrow

$$\langle f, g \rangle : \langle a, b \rangle \longrightarrow \langle a + b,\ a + b \rangle$$

Hence, we simply have to show that for any arrow $\langle f, g \rangle : \langle a, b \rangle \longrightarrow \langle c, c \rangle$, there is a morphism $f + g : a + b \longrightarrow c$. This becomes clear considering the following diagram.

[4] We use the definition of an adjunction as stated in Remark 12.3.4.

$$\begin{tikzcd}
\langle a,b\rangle \arrow[r, "\langle in_l, in_r\rangle"] \arrow[dr, "\langle f,g\rangle"'] & \langle a+b,a+b\rangle \arrow[d, "\langle f+g,f+g\rangle"] \\
 & \langle c,c\rangle
\end{tikzcd}$$

Given $\langle f,g\rangle : \langle a,b\rangle \longrightarrow \langle c,c\rangle$, the uniqueness of

$$\langle f+g,f+g\rangle : \langle a+b,a+b\rangle \longrightarrow \langle c,c\rangle$$

follows from the uniqueness of the coproduct. This can be easily checked. It is clear that the above diagram commutes.

Because of duality reasons, products can similarly defined using adjunctions. This time products are right adjoints of the functor G defined above.

12.4.3 Data Structures

We want to describe a very elementary data structure as an example for an adjunction.[5] We can recursively define functions on lists as follows:

$\texttt{length}(nil) = 0$
$\texttt{length}(s\diamondsuit t) = \texttt{length}(s) + \texttt{length}(t)$
$\texttt{length}(unit(x)) = 1$

The defined function `length` takes as input a list s and maps this list to a natural number specifying the length of s. In a similar way, we can recursively define a function that collects all elements of a list set theoretically:

$\texttt{element}(nil) = \emptyset$
$\texttt{element}(s\diamondsuit t) = \texttt{element}(s) \cup \texttt{element}(t)$
$\texttt{element}(unit(x)) = \{x\}$

We can interpret this situation in category theoretic terms as follows. Consider the two monoids[6] `lists` and `numbers` defined as follows:

$\texttt{lists} = \langle \texttt{list}(S), \diamondsuit, nil\rangle$
$\texttt{numbers} = \langle \mathbb{N}, +, 0\rangle$

Now consider the following diagram:

[5] The following example is due to [Ry86].

[6] Monoids are structures with an operation +, such that + is associative and there is a zero-element 0 with the property: $x + 0 = 0 + x = x$.

$$\begin{tikzcd}
S \arrow[r, "\texttt{unit}"] \arrow[dr, "f"'] & \texttt{List}(S) \arrow[d, "f^\sharp = \texttt{length}"] \\
& \mathbb{N}
\end{tikzcd}$$

In the above diagram, f is a function that takes any element of S to the number of elements of that list. Notice that $f^\sharp$ is a homomorphism. Moreover, $f^\sharp$ defines the length of a list. The diagram above describes a situation that is quite close to an adjunction, although no functors are defined. We can do this quite easily by a defining a functor F mapping a list to itself and a functor G mapping this list to the singleton set of that list. Then, `unit` is a natural transformation and in total the requirements of an adjunction are satisfied.

The same procedure yields an adjunction in the case of the definition of a function collecting elements of a list. Here the monoid $\langle \mathbb{N}, +, 0\rangle$ must be replaced by $\langle\langle Set_{Fin}(S), \cup, \emptyset\rangle$. Furthermore, some other trivial adjustments are necessary, but the outcome is the same. Again we can view the definition of that function collecting elements of a list as an adjunction.

The last section in this chapter about category theory summarizes facts concerning final objects of categories. Final objects (in particular final coalgebras) play an important role in the development of non-well-founded set theory and the applications in situation theory.

12.5 Final Objects

There is a further category theoretic concept we need to introduce. This concept describes the idea that certain objects of a given category can be designated as objects with special properties. Assume a category $\mathcal{C}$ is given. Then we want to know whether there is an object $a \in Obj_{\mathcal{C}}$, such that for every other object $b \in Obj_{\mathcal{C}}$ there is precisely one arrow $f : b \longrightarrow a$. This designated object is called final object. On the other hand, we want to examine the existence and the properties of a dual object, namely an object $c \in Obj_{\mathcal{C}}$, such that for every object $d \in Obj_{\mathcal{C}}$ there is precisely one arrow $g : c \longrightarrow d$. The latter object is called an initial object. Although the idea is simple we state the explicit definition here.

Definition 12.5.1 *(i) Assume $\mathcal{C}$ is a given category. An object $a \in Obj_{\mathcal{C}}$ is called initial, if for every $b \in Obj_{\mathcal{C}}$ there exists a unique arrow $f \in Ar_{\mathcal{C}}$, such that $f : b \longrightarrow a$.*

(ii) Assume again a category $\mathcal{C}$ is given. An object $c \in Obj_{\mathcal{C}}$ is called final, if for every object $d \in Obj_{\mathcal{C}}$ there exists a unique arrow $g \in Ar_{\mathcal{C}}$, such that $g : d \longrightarrow c$.

Although the idea of the concept of an isomorphism in category theory is the same as in other algebraic theories, it is useful to state the definition here explicitly.

Definition 12.5.2 *Assume a category $\mathcal{C}$ is given. An arrow $f : a \longrightarrow b$ is an isomorphism if and only if there is an arrow $g : b \longrightarrow a$, such that $g \circ f = id_a$ and $f \circ g = id_b$.*

We give some examples for initial and terminal objects in order to make the reader more familiar with these abstract concepts.

Example 12.5.1 (i) Consider the category $\mathcal{SET}$. As we explained in Example 12.1.3(ii) the arrows are arbitrary functions form sets into sets. Because the objects of $\mathcal{SET}$ do not have an algebraic structure (like groups or topological spaces do have) two objects are isomorphic if and only if they have the same cardinality. This implies that every object is final in $\mathcal{SET}$, if this object is precisely a one-element set. That holds because from every set a there is precisely one function mapping a into the one-element set. Notice that all sets with the same cardinality are isomorphic. Therefore, the final object in $\mathcal{SET}$ is unique up to isomorphisms. Concerning initial objects, the empty set $\emptyset$ is an initial object, because for every object $a \in Obj_{\mathcal{SET}}$ there is a unique $f : \emptyset \longrightarrow a$. Notice that $f = \{\emptyset\}$.

(ii) Consider the category $\mathcal{REC}$ of simple recursion data. $\mathcal{REC}$ is specified as follows: the objects are triples $\langle X, x, f\rangle$ where X is a set, $x \in X$, and $f : X \longrightarrow X$ is a (total) function. An arrow $\phi : \langle X, x, f\rangle \longrightarrow \langle Y, y, g\rangle$ is a function $\phi : X \longrightarrow Y$, such that $\phi(x) = y$ and $\phi(f(x)) = g(\phi(x))$. Identities are clear and composition is ordinary composition of functions. What is the initial object of $\mathcal{REC}$? Consider the triple $\langle \mathbb{N}, 0, s\rangle$, i.e. the natural numbers with 0 and the standard successor function. Obviously, there is a unique arrow f from the triple $\langle \mathbb{N}, 0, s\rangle$ to every other object of $\mathcal{REC}$. Hence, the structure $\langle \mathbb{N}, 0, s\rangle$ is the initial object of $\mathcal{REC}$. Concerning final objects it turns out that $\mathcal{REC}$ has also final objects. Final objects are triples $\langle \{x\}, \emptyset, id\rangle$ where $\{x\}$ is an arbitrary one-element set.

(iii) Consider a partially ordered set $\mathbf{D} = \langle D, \leq\rangle$ as a category. In general, $\mathbf{D}$ has no initial or final objects. This changes if $\mathbf{D}$ is a CCPO. Then the bottom element $\perp$ of $\mathbf{D}$ is initial because there is precisely one arrow to all other points $d \in D$. If $\mathbf{D}$ has also a top element $\top$, then $\top$ is final, because the resulting structure is a complete lattice. Notice that every complete lattice $\mathbf{D} = \langle D, \leq\rangle$ has a final and initial object. Furthermore, every complete semilattice has an initial object.

As we saw in Example 12.5.1(i), there are infinitely many final objects in $\mathcal{SET}$, but all these objects are isomorphic. We can generalize this fact as follows: in every category $\mathcal{C}$ final and initial objects are unique up to isomorphisms provided they exist.

Proposition 12.5.3 *Assume $\mathcal{C}$ is a category with final and initial objects. Then it holds: the initial and final objects are unique up to isomorphisms.*

Proof: Assume $\mathcal{C}$ is a category with initial and final objects. Assume further that a and b are two final objects. Notice that id_a and id_b are unique. Therefore, the following diagram commutes.

$$\begin{tikzcd}
a \arrow[r, "!f"] \arrow[d, "!id_a"'] & b \arrow[d, "!id_b"] \arrow[dl, "!g"'] \\
a & b \arrow[l, "!f"]
\end{tikzcd}$$

Obviously, all arrows in the diagram are unique. Using Definition 12.5.2 it follows that a and b are isomorphic. The proof concerning initial objects is the dual of the above reasoning. q.e.d.

An important concept for applications is the correspondence between fixed points and initial or final objects. We will prove the following correspondences: the initial object a of a given category can be associated with a fixed point of a certain endofunctor Γ. The same is true for the final object. In general, the two fixed points are different because the final and initial objects are different in general. The following facts make that intuition precise.

Fact 12.5.4 *Assume a category $\mathcal{C}$ and an endofunctor Γ is given, such that for every object $\Gamma(b) \in Obj_{\mathcal{C}}$ there is a morphism $f : b \longrightarrow \Gamma(b)$. If an object $a \in Obj_{\mathcal{C}}$ is final, then a is a fixed point of Γ.*

Proof: Let a be a final object in $\mathcal{C}$. Now, consider the object $b = \Gamma(a)$. By assumption there is a morphism $f : a \longrightarrow \Gamma(a)$. Then the following holds:

$$\Gamma(a) \xrightarrow{f} a \xrightarrow{g} \Gamma(a) \xrightarrow{f} a$$

By uniqueness (up to isomorphisms) of the identity morphism id_a and $id_{\Gamma(a)}$ it follows that $f \circ g = id_{\Gamma(a)}$ and $g \circ f = id_a$. This shows that f is the inverse morphism of g and vice verse. Conclude: a is a fixed point (up to isomorphisms). q.e.d.

The following fact shows that the dual statement is also true: every initial object is a fixed point of an endofunctor Γ.

Fact 12.5.5 *Assume a category $\mathcal{C}$ and an endofunctor Γ is given, such that for every object $b \in Obj_{\mathcal{C}}$ there is a morphism $f : \Gamma(b) \longrightarrow b$. If an object $a \in Obj_{\mathcal{C}}$ is initial, then a is a fixed point of Γ.*

Proof: Dual of Fact 12.5.4 q.e.d.

Some further remarks are added at the end of this chapter.

Remark 12.5.2 (i) Intuitively, one associates with a maximal fixed point a final object and with a minimal fixed point a initial object. The problem is that without further specifications of the category $\mathcal{C}$ one has no order relation where the notions maximal vs. minimal fixed point can refer. For special cases this is less problematic as we will see later.

(ii) It is illuminating to consider the additional conditions on Γ in Facts 12.5.4 and 12.5.5 in an easy example. If one considers $\mathbf{D} = \langle D, \leq \rangle$ as a category, then the existence of a morphism $f : b \longrightarrow \Gamma(b)$ translates into the condition: $\forall b \in D : b \leq \Gamma(b)$, whereas the condition on Γ in Fact 12.5.5 translates into $\forall b \in D : \Gamma(b) \leq b$. In the first case, this property of Γ is usually called inductive.

(iii) Facts 12.5.4 and 12.5.5 give a first flavor of the situation in the theory of coalgebras as described in the next chapter. Here, an endofunctor $\Gamma : \mathcal{C} \longrightarrow \mathcal{C}$ is essentially embedded in the definition of a coalgebra. A similar result holds also in the case of coalgebras.

We finish the introduction into category theory with these remarks. More could be said about further types of categories. For further information concerning general concepts in category theory the reader is referred to the standard literature mentioned in the History section.

12.6 History

There are many good introductions in category theory. The standard textbook is still [Ma71]. A readable introduction is [Sc70] (or the English version [Sc72]). Another alternative is [Pa70]. Good tutorials for quite different concepts of category theory and a good resource of many applications for computer science can be found in the book [GoHa86]. In particular, this text as well as the text [BaWe90] are valuable for someone who is especially interested in using category theory in the theory of programming or in computer science in general. Our informal presentation of category theory is in the spirit of [Sc70]. This book is translated from German into English, revised and slightly enlarged in the English edition.

The origins of category theory go back to the late 50s motivated by topology. The invention of the theory is usually prescribed to MacLane and Eilenberg referring to their article [EiMa45]. After that time, a rapid development took place, first with respect to applications in topology, but later because of the fact that category theory was successfully used in different algebraic theories.

The idea to use category theory in logic is not very old. The classical field where category theory initiated new results in logic is linear logic. Most prominently it was Girard who developed this application of category theory. With the help of category theory Girard was able to give the first model for linear logic. Classical resources for more information about term categories and the usage of category theory in substructural logic are [Da92, Tr92], and [GiLaTa89].

Chapter 13

Coalgebras and Transition Systems

In the last chapter, we introduced a lot of category theoretic concepts. Although category theory is a topic that justifies interest and research in its own right, our primary concern is to use these techniques for circular phenomena in this chapter and the following ones. First, we will introduce some basic concepts of theoretical computer science and the theory of coalgebras. Second, we will model labeled transition systems and streams in category theoretic terms. After that we will examine the principle of proof by coinduction in a wider context. The main aim of this chapter is to present a possibility to model certain concepts in theoretical computer science. In particular, we will emphasize the theory of deterministic labeled transition systems as the most prominent example. Last but not least, we will associate systems of equations with the theory of coalgebras.

13.1 Coalgebras and Some Properties

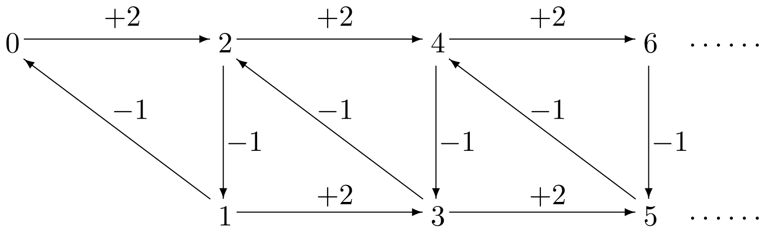
We begin our consideration with labeled transition systems. Labeled transition systems can be found in various fields in theoretical computer science and are a generalization of the theory of finite state automata. Contrary to finite state automata labeled transition systems can be infinite. We want to discover how labeled transition systems can be represented as Γ-coalgebras.

Definition 13.1.1 *Let A be an arbitrary set (of labels) and S be a set of states. The triple $\langle S, \longrightarrow_S, A\rangle$ is called a labeled transition system if it holds: $\longrightarrow_S \subseteq S \times A \times S$. We call $\longrightarrow_S$ the transition relation (or the dynamics) of the labeled transition system $\langle S, \longrightarrow_S, A\rangle$.*

Example 13.1.1 Assume $A = \{+2, -1\}$ and the set of states is equivalent to the set of the natural numbers: $S = \mathbb{N} = \{0, 1, 2, ...\}$. We define a labeled transition system $\langle S, \longrightarrow_S, A\rangle$ for the defined sets S and A via the transition relation $\longrightarrow_S$ as follows:

$$\langle s, a, s' \rangle \in \longrightarrow_S \iff s' = s + 2 \text{ or } s' = s - 1$$

The considered labeled transition system $\langle S, \longrightarrow_S, A \rangle$ can be represented as an infinite graph as follows:



The next definition provides the main concept for this (and the following) sections. It is the concept of a coalgebra. Although coalgebras are formulated as an abstract concept that was not connected with the theory of automata originally, coalgebras can be considered as a generalization of labeled transition systems.

Definition 13.1.2 *Let $\mathcal{SET}$ be the category of all sets, $\Gamma : \mathcal{SET} \longrightarrow \mathcal{SET}$ be a functor, and S be an arbitrary set. A pair $\langle S, \alpha_S \rangle$ is called a Γ-coalgebra (also called Γ-system), if $\alpha_S : S \longrightarrow \Gamma(S)$. We call S the carrier or set of states of the Γ-coalgebra and α_S the Γ-transition structure (also called the dynamics of the system).*

Remark 13.1.2 (i) Our definition of a Γ-coalgebra is completely general with no restrictions on the functor Γ. Later we will consider Γ-coalgebras with additional restrictions concerning Γ. Most prominently we will consider coalgebras where Γ is composed of certain simple operations.

(ii) In Definition 13.1.2, Γ is a functor mapping a specified category $\mathcal{SET}$ into itself. An obvious generalization of Definition 13.1.2 would be to drop this specification. We can say that given an arbitrary category $\mathcal{C}$ the functor Γ has only to satisfy the condition to be an endofunctor. This is sufficient to define Γ-coalgebras. For the purpose of this chapter it is not necessary to be as general as possible.

The next fact points out the close connection between labeled transition systems and Γ-coalgebras. It turns out that labeled transition systems are precisely Γ-coalgebras for a particular functor Γ.

Fact 13.1.3 *Every labeled transition system can be interpreted as a Γ-coalgebra and every Γ-coalgebra with $\Gamma : S \longrightarrow A \times S$ can be interpreted as a labeled transition system.*

Proof: "$\Rightarrow$" Assume $\langle S, \longrightarrow_S, A\rangle$ is a labeled transition system. In order to find a representation in terms of a Γ-coalgebra, we need to characterize Γ itself. Define Γ according to the following condition:

$$\Gamma(S) \ = \ \{Y \mid Y \subseteq A \times S\} \ = \ \wp(A \times S)$$

We can choose for the Γ-transition structure the following condition, where we mirror precisely the situation in labeled transition systems:

$$\alpha_S : S \longrightarrow \Gamma(S) : s \longmapsto \{\langle a, s'\rangle \mid a \in A \ \wedge \ s' \in S \ \wedge \ \langle s, a, s'\rangle \in \longrightarrow_S\}$$

The Γ-coalgebra $\langle S, \alpha_S\rangle$ is an equivalent representation of the labeled transition system $\langle S, \longrightarrow_S, A\rangle$.

"$\Leftarrow$" Assume that $\langle X, \alpha_X\rangle$ is a Γ-coalgebra for the fixed functor $\Gamma : X \longrightarrow A \times X$. Consider the triple $\langle X, \longrightarrow_X, A\rangle$, with the following identification

$$\langle x, a, x'\rangle \ \in \ \longrightarrow_X \ \Longleftrightarrow \ \langle a, x'\rangle \in \alpha_X(x)$$

Then, the labeled transition system $\langle X, \longrightarrow_X, A\rangle$ represents the Γ-coalgebra $\langle X, \alpha_X\rangle$. This suffices to shows the fact. q.e.d.

We want to define structure preserving relations between coalgebras, like homomorphisms between groups, continuous functions between topological spaces, or order preserving mappings between partially ordered sets. In order to do this, we introduce the category of all coalgebras $\mathcal{CO}$. We will see that homomorphisms between coalgebras are precisely the arrows of $\mathcal{CO}$. Arrows in $\mathcal{CO}$ can be considered as structure preserving mappings between coalgebras.

Definition 13.1.4 *Let $\Gamma : \mathcal{C} \longrightarrow \mathcal{C}$ be an endofunctor defined on a category $\mathcal{C}$. The category $\mathcal{CO}$ of all Γ-coalgebras is defined as follows.*

- *The objects are Γ-coalgebras $\langle S, \alpha_S\rangle$.*
- *The arrows of $\mathcal{CO}$ are homomorphisms $f : \langle S, \alpha_S\rangle \longrightarrow \langle T, \alpha_T\rangle$, such that $\Gamma(f) \circ \alpha_S = \alpha_T \circ f$.*
- *Concatenation $\circ$ is clear. (It is also clear that concatenation of arrows is associative and that the concatenation of two arrows is again an arrow.)*
- *The domain dom and the codomain cod of an arrow f are clear.*
- *The identity arrow for a Γ-coalgebra $\langle S, \alpha_S\rangle$ is an arrow $id_S : S \longrightarrow S$, such that $\Gamma(id_S) \circ \alpha_S = \alpha_S \circ id_S$. It is clear that such an arrow exists for every Γ-coalgebra.*

In the category $\mathcal{CO}$, the arrows (homomorphisms) are functions from Γ-coalgebras into Γ-coalgebras, such that Γ is compatible with the transitions structure of the Γ-coalgebras. In other words, arrows respect transitions and reflections of the dynamics of the coalgebra. They preserve the structure in both direction.

Definition 13.1.5 *Assume two Γ-coalgebras $\langle S, \alpha_S \rangle$ and $\langle T, \alpha_T \rangle$ are given where $\Gamma : \mathcal{SET} \longrightarrow \mathcal{SET}$. A homomorphism $f : S \longrightarrow T$ is a function, such that $\Gamma(f) \circ \alpha_S = \alpha_T \circ f$. In other words: a homomorphism $f : S \longrightarrow T$ is an arrow in $\mathcal{CO}$.*

Remark 13.1.3 (i) Definition 13.1.5 expresses essentially that a homomorphism f preserves the transition structure of the coalgebra $\langle S, \alpha_S \rangle$ if mapped to a coalgebra $\langle T, \alpha_T \rangle$ in both directions. If one takes into account that Definition 13.1.5 is literally the same as the definition of arrows in the category $\mathcal{CO}$ the concept homomorphism is quite natural.
The diagrammatic representation of the definition of a homomorphism is captured in the following diagram. Assume as in Definition 13.1.5 that two Γ-coalgebras $\langle S, \alpha_S \rangle$ and $\langle T, \alpha_T \rangle$ are given. A mapping $f : S \longrightarrow T$ is a homomorphism, if the following diagram commutes:

$$\begin{tikzcd}
S \arrow[r, "f"] \arrow[d, "\alpha_S"'] & T \arrow[d, "\alpha_T"] \\
\Gamma(S) \arrow[r, "\Gamma(f)"'] & \Gamma(T)
\end{tikzcd}$$

(ii) As usual we define an injective homomorphism as a monomorphism and a surjective homomorphism as an epimorphism. Similarly to the situation in classical algebra, every injective and surjective homomorphism is an isomorphism. That follows directly from the definition of an isomorphism.

(iii) It is easy to check that in the theory of labeled transition systems homomorphisms can be defined by the following conditions (a) and (b):

(a) $(\forall s \in S)(\forall a \in A) : (s \xrightarrow{a} s') \implies (\Gamma(s) \xrightarrow{a} \Gamma(s'))$
(b) $(\forall s \in S)(\forall a \in A) : (\Gamma(s) \xrightarrow{a} t') \implies (\exists s' : s \xrightarrow{a} s')$

Whereas (a) is a property preserving the transition structure, (b) is called the reflecting property. It is clear that (a) and (b) together are equivalent to the requirements in Definition 13.1.4 of arrows in the category $\mathcal{CO}$.

We need a possibility to determine whether two given coalgebras are structurally identical or not. In algebraic theories, one uses isomorphisms in order to prove that two structures are equal. In the coalgebraic framework, this is no longer sufficient. Therefore, one introduces bisimulations in order to check structural similarity. The formal definition of a bisimulation can be formulated as follows.

Definition 13.1.6 *Assume $\Gamma : \mathcal{SET} \longrightarrow \mathcal{SET}$ is a given functor. Assume further that $\langle S, \alpha_S \rangle$ and $\langle T, \alpha_T \rangle$ are two Γ-coalgebras. A Γ-bisimulation between $\langle S, \alpha_S \rangle$ and $\langle T, \alpha_T \rangle$ is a coalgebra $\langle R, \alpha_R \rangle$ such that $R \subseteq S \times T$ and the following diagram commutes:*

$$\begin{tikzcd}
S \arrow[d, "\alpha_S"'] & R \arrow[l, "\pi_1"'] \arrow[r, "\pi_2"] \arrow[d, "\alpha_R"] & T \arrow[d, "\alpha_T"] \\
\Gamma(S) & \Gamma(R) \arrow[l, "\Gamma(\pi_1)"] \arrow[r, "\Gamma(\pi_2)"'] & \Gamma(T)
\end{tikzcd}$$

We say that for $s \in S$ and $t \in T$ state s is bisimilar to state t, if there is a bisimulation $\langle R, \alpha_R \rangle$, such that $\langle s, t \rangle \in R$.

Example 13.1.4 Assume $\langle S, \alpha_S \rangle$ and $\langle T, \alpha_T \rangle$ are two labeled transition systems. A subset $R \subseteq S \times T$ is a bisimulation between $\langle S, \alpha_S \rangle$ and $\langle T, \alpha_T \rangle$ if the following two conditions hold:

- If $\langle s, t \rangle \in R$, then it holds: $\forall s' \in S : s \xrightarrow{a} s' \;\Rightarrow\; \exists t' \in T : t \xrightarrow{a} t'$ and $\langle s', t' \rangle \in R$
- If $\langle s, t \rangle \in R$, then it holds: $\forall t' \in T : t \xrightarrow{a} t' \;\Rightarrow\; \exists s' \in S : s \xrightarrow{a} s'$ and $\langle s', t' \rangle \in R$

Notice that this is the familiar definition of a bisimulation: everything that can be reached in $\langle S, \alpha_S \rangle$ has a counterpart with the same properties in $\langle T, \alpha_T \rangle$ and vice versa.

We show in the following fact that Definition 13.1.6 and the representation of a bisimulation in Example 13.1.4 are equivalent.

Fact 13.1.7 *Definition 13.1.6 of a bisimulation is equivalent for labeled transition systems with the conditions:*

- *If $\langle s, t \rangle \in R$, then it holds: $\forall s' \in S : s \xrightarrow{a} s' \;\Rightarrow\; \exists t' \in T : t \xrightarrow{a} t'$ and $\langle s', t' \rangle \in R$*
- *If $\langle s, t \rangle \in R$, then it holds: $\forall t' \in T : t \xrightarrow{a} t' \;\Rightarrow\; \exists s' \in S : s \xrightarrow{a} s'$ and $\langle s', t' \rangle \in R$*

Proof: For one direction assume that $R \subseteq S \times T$ is a bisimulation according to the above conditions. We can define a transition structure on R as follows: $\alpha_R : R \longrightarrow \wp(R)$, such that the following condition holds:

$$\alpha_R(\langle s,t\rangle) = \{\langle s',t'\rangle \mid \langle s',t'\rangle \in R \ \wedge \ s \xrightarrow{a} s' \ \wedge \ t \xrightarrow{a} t'\}$$

Obviousely, the projection $\pi_1 : R \longrightarrow S$ and $\pi_2 : R \longrightarrow T$ are homomorphisms.

For the other direction assume (according to Definition 13.1.6) that $R \subseteq S \times T$, and $\alpha_R : R \longrightarrow \Gamma(R)$. Additionally, assume that the corresponding diagram commutes. As we saw in Fact 13.1.3, α_R induces a relation $\longrightarrow_R \subseteq R \times A \times R$. Assume that it holds $\langle s,t\rangle \in R$ and $s \xrightarrow{a} s'$. Because $\pi_1(\langle s,t\rangle) = s$ we have $\pi_1(\langle s,t\rangle) \xrightarrow{a} s'$. Because π_1 is a homomorphism there exists a pair $\langle s'',t'\rangle \in R$, such that $\pi_1(\langle s'',t'\rangle) = s'$ and $\langle s,t\rangle \xrightarrow{a} \langle s'',t'\rangle$. Conclude that $\langle s',t'\rangle \in R$. With the assumption that π_2 is a homomorphism we can conclude that $t \xrightarrow{a} t'$. This justifies the first condition. The second condition is proven similarly. q.e.d.

The following fact formulates an interesting correlation between homomorphisms and bisimulations. It turns out that the graph of a homomorphism is a bisimulation.

Fact 13.1.8 *Assume $\langle S,\alpha_S\rangle$ and $\langle T,\alpha_T\rangle$ are two Γ-coalgebras and $\Gamma : \mathcal{SET} \longrightarrow \mathcal{SET}$ is a given functor. Then it holds: A function $f : S \longrightarrow T$ is a homomorphism if and only if the graph $G(f)$ of f is a bisimulation between $\langle S,\alpha_S\rangle$ and $\langle T,\alpha_T\rangle$.*

Proof: For the left to right direction suppose that $f : S \longrightarrow T$ is a homomorphism. Consider the coalgebra $\langle G(f), \Gamma(\pi_1)^{-1} \circ \alpha_S \circ \pi_1\rangle$. We have to show that $\pi_1 : G(f) \longrightarrow S$ and $\pi_2 : G(f) \longrightarrow T$ are homomorphisms. First, consider π_1. The following equalities hold:

$$\begin{aligned}
\Gamma(\pi_1) \circ (\Gamma(\pi_1)^{-1} \circ \alpha_S \circ \pi_1) &= \Gamma(\pi_1 \circ \pi_1^{-1}) \circ \alpha_S \circ \pi_1 \\
&= \Gamma(id_R) \circ \alpha_S \circ \pi_1 \\
&= \alpha_S \circ \pi_1
\end{aligned}$$

Therefore, π_1 is a homomorphism. Concerning π_2, we can state the following equalities.

$$\begin{aligned}
\Gamma(\pi_2) \circ (\Gamma(\pi_1)^{-1} \circ \alpha_S \circ \pi_1) &= \Gamma(\pi_2 \circ \pi_1^{-1}) \circ \alpha_S \circ \pi_1 \\
&= \Gamma(f) \circ \alpha_S \circ \pi_1 \\
&= \alpha_T \circ f \circ \pi_1 \\
&= \alpha_T \circ \pi_2
\end{aligned}$$

This suffices to show one direction of the theorem.

For the right to left direction assume that $\langle G(f), \alpha_{G(f)}\rangle$ is a bisimulation between $\langle S,\alpha_S\rangle$ and $\langle T,\alpha_T\rangle$. We define $f := \pi_1^{-1} \circ \pi_2$ where π_1 and π_2 are the well-known projection functions. We show that f is a homomorphism. Notice that f is well-defined, because π_1 is a bijective function and therefore the inverse function of π_1 exists and is a homomorphism. With the fact that concatenation of homomorphisms yields again a homomorphism (we are working in the category of coalgebras!) it is clear that $f := \pi_1^{-1} \circ \pi_2$ is a homomorphism. This completes the proof of the theorem. q.e.d.

What can be said about non-deterministic labeled transition systems concerning final systems? As was shown in Fact 13.1.3 non-deterministic labeled transition systems can be represented as Γ-coalgebras, if we choose the functor $\Gamma : S \longrightarrow \wp(A \times S)$. Unfortunately, for this functor Γ there does not exist a final coalgebra: the unbounded power set operation prevents the existence of final coalgebras. In order to see this, assume there was a final coalgebra $\langle \Gamma^*, \alpha_{\Gamma^*} \rangle$ for the power set functor $\wp$. Using the fact that the transition function $\alpha_{\Gamma^*} : \Gamma^* \longrightarrow \Gamma(\Gamma^*)$ is an isomorphism, we can conclude that it must hold: $\alpha_{\Gamma^*} : \wp^* \longrightarrow \wp(\wp^*)$ is an isomorphism. In other words, it holds: $\wp(\wp^*) \subseteq \wp^*$, which contradicts Cantors theorem, because there is no set $\wp^*$ that satisfies this condition. Conclude: there is no final Γ-coalgebra if Γ includes the unrestricted power set functor.[1]

There is a possibility to circumvent this problem by restricting $\wp^*$ appropriately. Consider the functor $\wp_{fin} : S \longrightarrow \{X \mid X \subseteq S \ \wedge \ |X| < \omega\}$. The claim is that the category of all $\wp_{fin}$-coalgebras has a final object. In order to state this result, we need a definition.

Definition 13.1.9 *Assume Γ is a functor from the category of all sets into itself. We call Γ a bound functor if there is a set V, such that for every coalgebra $\langle S, \alpha_S \rangle$ and for every $s \in S$ there exists an injective mapping f from the carrier of the sub-coalgebra $\langle s \rangle$ into V where the coalgebra generated by s and represented by $\langle s \rangle$ is defined as follows:*

$$\langle s \rangle \ := \ \bigcap \{ \langle X, \alpha_X \rangle \mid \langle X, \alpha_X \rangle \textit{ is sub-coalgebra of } \langle S, \alpha_S \rangle \textit{ and } s \in X \}$$

Using Definition 13.1.9 we can state the theorem claiming the existence of a final coalgebra for the restricted power set operation $\wp_{fin}$.

Theorem 13.1.10 *There exists a final coalgebra for the functor $\wp_{fin} : S \longrightarrow \{X \mid X \subseteq S \wedge |X| < \omega\}$.*

Proof: First, we prove that for $\wp_{fin}$ there exists a set of generators. If $\langle S, \alpha_S \rangle$ is a $\wp_{fin}$-coalgebra and $s \in S$ is arbitrarily chosen, then for every natural number $n \in \mathbb{N}$, there is only a finite number of states that can be reached from s. Consider for s the generated coalgebra

$$\langle s \rangle \ = \ \bigcap \{ \langle V, \alpha_V \rangle \mid \langle V, \alpha_V \rangle \text{ is a sub-coalgebra of } \langle S, \alpha_S \rangle \text{ and } s \in V \}$$

[1] Notice that we are working in the category $\mathcal{SET}$. It is well-known that there is a class Γ^* that satisfies the condition $\wp(\wp^*) \subseteq \wp^*$. In the category $\mathcal{SET}$, this collection is not an object of $\mathcal{SET}$.

Notice that $\langle s\rangle$ has at most countably many states and therefore there is a partial isomorphism (therefore an injective mapping) from $\langle s\rangle$ into $\langle \mathbb{N}, 0,' \rangle$ (where $'$ is the successor function and $\mathbb{N}$ is the set of natural numbers). This shows that $\wp_{fin}$ is bound. Using this fact one can define a final coalgebra by taking the coproduct of all generated coalgebras for all $\wp_{fin}$-coalgebras. In other words, we know that it holds:

$$(\forall\langle S,\alpha_S\rangle)(\forall s\in S)(\exists\langle G_i,\alpha_i\rangle) : \langle S,\alpha_S\rangle \cong \langle G_i,\alpha_i\rangle$$

In the above formula, $\langle G_i,\alpha_i\rangle$ is the generated coalgebra $\langle s\rangle$. Consider the coproduct $\langle U,\alpha_U\rangle$ of all generated coalgebras $\langle G_i,\alpha_i\rangle$.

$$\langle U,\alpha_U\rangle = \coprod\{\langle G_i,\alpha_i\rangle \mid \langle S,\alpha_S\rangle \cong \langle G_i,\alpha_i\rangle\}$$

Because of the remarks above there exists an injective homomorphism g from the coproduct into each coalgebra $\langle S,\alpha_S\rangle$:

$$g : \langle U,\alpha_U\rangle \longrightarrow \langle S,\alpha_S\rangle$$

.
Define the final coalgebra as the quotient algebra $\langle P,\pi\rangle = \langle U/_{\sim U}, b_{\sim_U}\rangle$ where $\sim_U$ is the maximal bisimulation on U. It remains to show that $\langle P,\pi\rangle$ really is a final coalgebra.

In order to show the claim, notice the following fact: For $f : S \longrightarrow U$ and $f' : U \longrightarrow P$ where f is a partial isomorphism and f' is a quotient homomorphism the composition $h = f' \circ f$ is again a homomorphism (concatenation of morphisms in a category). That means that for every coalgebra $\langle S,\alpha_S\rangle$ there exists a homomorphism $h : S \longrightarrow P$. We have to show that h is unique. First, notice that $\langle P,\pi\rangle$ has no proper quotients. To see this, consider a mapping $g : P \longrightarrow P/_{\sim_U}$. Obviously, g is an isomorphism. Then, $\sim_U$ is the maximal bisimulation, i.e. for every bisimulation R it holds: $R \subseteq \sim_U$. Last but not least, assume that $h : S \longrightarrow P$ and $h' : S \longrightarrow P$ are two homomorphisms. Consider the relation Q specified as follows:

$$Q = \{\langle G(h), G(h')\rangle \mid G(h) \text{ and } G(h') \text{ are graphs of } h \text{ and } h'\}$$

Then Q is a bisimulation on P, because of the well-known fact that concatenation of bisimulations are again bisimulations. Because $\sim_U$ is maximal, we have $Q \subseteq \sim_U$ and therefore $h = h'$. Conclude: the homomorphism $h = f' \circ f$ is unique. q.e.d.

We add an important additional fact concerning final coalgebras. Every final coalgebra in a given category $\mathcal{C}$ is a fixed point of the endofunctor $\Gamma : \mathcal{C} \longrightarrow \mathcal{C}$. This fact corresponds closely to Fact 12.5.4. We state this fact as a proposition.

Proposition 13.1.11 *Assume $\mathcal{C}$ is a category of coalgebras and $\Gamma : \mathcal{C} \longrightarrow \mathcal{C}$ is an endofunctor. If the coalgebra $\langle P,\pi\rangle$ is final, then $\langle P,\pi\rangle$ is a fixed point of Γ.*

Proof: Assume $\langle P, \pi \rangle$ is a final coalgebra for the functor $\Gamma : \mathcal{C} \longrightarrow \mathcal{C}$. Clearly, the pair $\langle \Gamma(P), \Gamma(\pi) \rangle$ is again a $\Gamma-$coalgebra. Consider the following diagram.

$$\begin{tikzcd}
P \arrow[r, "f"] \arrow[d, "\pi"'] & \Gamma(P) \arrow[r, "\gamma"] \arrow[d, "\Gamma(f)"] & P \arrow[d, "\pi"] \\
\Gamma(P) \arrow[r, "\Gamma(f)"'] & \Gamma^2(P) \arrow[r, "\Gamma(\gamma)"'] & \Gamma(P)
\end{tikzcd}$$

Obviously, $f : P \longrightarrow \Gamma(P)$ is a morphism in $\mathcal{C}$. Because of the finality of $\langle P, \pi \rangle$ the morphism $\gamma \circ f : P \longrightarrow P$ is the identity on P. Then, the morphism $\Gamma(\gamma) \circ \Gamma(f)$ is the identity $id_{\Gamma(P)}$. Conclude $\langle P, \pi \rangle$ is a fixed point. q.e.d.

In the next section, we will examine how certain circular phenomena can be represented in the theory of deterministic labeled transition systems. We will use our coalgebraic framework developed so far in the following.

13.2 Properties of Labeled Transition Systems

In this section, we will have a closer look concerning certain phenomena of deterministic labeled transition systems. We will see that the objects we will examine are not circular in the sense that we need necessarily the theory of hypersets in order to model these objects, but they have strong similarities to circular phenomena. This is intuitively clear if one considers a labeled transition system that contains circles. In this and the following sections the term labeled transition systems refers to deterministic labeled transition systems.

Focusing on the theory we have developed so fare we can begin with an examination of maximal fixed points. Consider the functor $\Gamma : S \longrightarrow A \times S$. We would like to know what the final coalgebra of this functor is. It turns out that the coalgebra $\langle A^\omega, \langle h, t \rangle \rangle$ is final for Γ where A^ω is the collection of all infinite streams, h is a function mapping an arbitrary stream to its first argument, and t is a function mapping an arbitrary stream s to the stream s' resulting from s by deleting the first element. The following fact guarantees that A^ω is a maximal fixed point of Γ.

Fact 13.2.1 *For the functor $\Gamma : S \longrightarrow A \times S$, the set A^ω is a maximal fixed point, i.e. it holds $\Gamma(A^\omega) = A^\omega$ and for every fixed point X it holds: $X \subseteq A^\omega$.*

Proof: First, we show that A^ω is a fixed point, i.e. we show that it holds: $A^\omega = A \times A^\omega$. Assume $a = \langle a_1, a_2, \dots \rangle \in A^\omega$, then obviously the pair $\langle a_1, \langle a_2, a_3, \dots \rangle \rangle \in A \times A^\omega$. Up to isomorphisms it holds: $\langle a_1, \langle a_2, a_3, \dots \rangle \rangle = \langle a_1, a_2, \dots \rangle$. Hence, we have $A^\omega \subseteq A \times A^\omega$. Clearly, the other direction is also true. This shows that A^ω is a fixed point. It remains to show that A^ω is also maximal. Assume $X = A \times X$ is an arbitrary fixed point of Γ. Consider an arbitrary $x \in X$. Then, obviously we have $x \in A \times X$ because X is a fixed point. Hence, w.l.o.g. $x = \langle a_1, x_1 \rangle$, for an appropriate $x_1 \in X$. Applying this procedure one more time we get: $x = \langle a_1, a_2, x_2 \rangle$ where again $x_2 \in X$. Continuing this process yields the stream $x = \langle a_1, a_2, \dots \rangle$. Obviously, x is of the form $x \in A^\omega$. Conclude that the set of all infinite streams A^ω is maximal for Γ. q.e.d.

Now we want to associate Fact 13.2.1 with some ideas from the theory of non-well-founded set theory as developed in Chapter 11. Although we did not use the anti-foundation axiom in the proof of Fact 13.2.1 above, the claim that A^ω is a fixed point is very close to the statement that A^ω is the solution of a particular system of equations as developed in Chapter 11. We consider two examples in the following Remark 13.2.1.

Remark 13.2.1 (i) Consider the system $\langle \{x\}, \{a\}, e \rangle$ where $x \in \mathcal{U}$, $a \in \mathcal{U}$, $x \neq a$, and furthermore

$$e : x \longmapsto \{\langle a, x \rangle\} = \{\{\{a\}, \{a, x\}\}\}$$

We can apply the anti-foundation axiom for general systems and can establish a solution for the above system as follows:

$$s(x) = \{\{\{a\}, \{a, s(x)\}\}\}$$

Unfolding $s(x)$ shows that we get the stream $\langle a, a, a, \dots \rangle$ (up to isomorphisms identifying different codings of infinite sequences). Hence, the unfolded stream that is bisimilar to $s(x)$ is an element in A^ω.

(ii) We consider a slight change of the underlying system of equations. Assume a system is given by $\langle \{x, y\}, \{a\}, e \rangle$ with $\{x, y\} \subseteq \mathcal{U}$, $a \in \mathcal{U}$, and e defined as follows: $e(x) = \langle a, y \rangle$ and $e(y) = \langle a, x \rangle$, then we get the following solutions:

$$s(x) = \{\{\{a\}, \{a, s(y)\}\}\}$$

$$s(y) = \{\{\{a\}, \{a, s(x)\}\}\}$$

Notice that $s(x)$ and $s(y)$ are bisimilar in the sense of Definition 11.2.2. The two solutions determine the same stream $\langle a, a, \dots \rangle \in A^\omega$.

Although the above type of system fits in the framework of non-well-founded set theory developed so far, one restriction is that we are forced to require that $A \subseteq \mathcal{U}$. The following definition of systems of equations gives not only a further representation of deterministic labeled transition systems, but also generalizes the above systems in the sense that sets can be elements of A.

Definition 13.2.2 *Consider the triple $\langle X, A, e\rangle$. If $X \subseteq \mathcal{U}$, A is a set, such that it holds $X \cap support(A) = \emptyset$, and $e : X \longrightarrow A \times X$ is a function, then we call $\langle X, A, e\rangle$ a flat system of streams.*

Remark 13.2.2 (i) Notice that $A \times X$ is not a subset of $\wp(A \cup X)$. Therefore, Definition 13.2.2 does not specify a flat system of equations. It is not a general system of equations either, because A is not a subset of urelements, but merely an arbitrary set with the property $support(A) \cap X = \emptyset$.

(ii) We can state the following claim: Definition 13.2.2 determines precisely the same class of structure as the category of all Γ-coalgebras for the functor $\Gamma : S \longrightarrow A \times S$ where each $s \in S$ denotes a state and A is a fixed set of labels for nodes. Assume $\langle X, A, e\rangle$ is a system of streams, then fix a set of states S, such that there is a bijective mapping $f : S \longrightarrow X$ and take A to be the set of labels.[2] Consider the pair $\langle S, \alpha_S\rangle$ where α_S is defined as follows:

$$\forall s \in S : \alpha_S(s) = \langle a, s'\rangle \;\Leftrightarrow\; e(x) = \langle a, x'\rangle \;\wedge\; f(s) = \{x\} \;\wedge\; f(s') = \{x'\}$$

Obviously, this gives us what we want. On the other hand, assume $\langle S, \alpha_S\rangle$ is a Γ-coalgebra for the functor $\Gamma : S \longrightarrow A \times S$. Then, X is in a bijective correspondence with S and the obvious dynamics guarantees the equivalence. This suffices to show that the two definitions are coextensional.

Now, we state some facts concerning flat system of streams, in particular, we need to guarantee that solutions for this type of system exist. Furthermore, we focus on the close relationship between deterministic labeled transition systems (and their outputs) on the one hand and flat systems of streams on the other hand.

Proposition 13.2.3 *(i) The anti-foundation axiom implies that every flat system of streams has a unique solution.*

(ii) Assume $\mathcal{E} = \langle X, A, e\rangle$ is a flat system of streams. Then it holds: the solution set of $\langle X, A, e\rangle$, denoted by $sol(\mathcal{E})$, is a subset of the maximal fixed point of the operator $\Gamma : S \longrightarrow A \times S$.

(iii) For every subset $B \subseteq A^\omega$ of the maximal fixed point of the operator $\Gamma : S \longrightarrow A \times S$, there is a flat system of streams $\mathcal{E} = \langle X, A, e\rangle$, such that $B = sol(\mathcal{E})$.

Proof: (i) Every flat system of streams $\langle X, A, e\rangle$ with $e : X \longrightarrow A \times X$ is a general system of equations of the form $e : X \longrightarrow V_{afa}[X \cup support(A)]$. Because of Theorem 11.6.3 there is a unique solution of this system by the anti-foundation axiom.

[2] We assume that all states $s \in S$ are atomic and are not further structured.

(ii) Assume $\langle X, A, e\rangle$ is a flat system of streams. Assume further that $sol(\mathcal{E})$ is the corresponding solution set. We have to show that $sol(\mathcal{E}) \subseteq A^\omega$. Consider the function $e : X \longrightarrow A \times X$, such that every $x \in X$ is mapped to a pair $\langle a, y\rangle$ where $y \in X$. The solution for $x \in X$ is given by the following substitution:

$$s(x) = \langle s(a), s(y)\rangle = \langle a, s(y)\rangle \in A \times sol(\mathcal{E})$$

Hence, we get the following relation:

$$sol(\mathcal{E}) \subseteq A \times sol(\mathcal{E})$$

This shows that the solution set of $\langle X, A, e\rangle$ is a set of streams. In order to show that every stream is infinite, notice that a termination of the 'dynamics' is not given: There is no $x \in X$, such that $s(x) = \langle a, b\rangle$, for $b \notin X$.

(iii) Assume $B \subseteq A^\omega$ is arbitrarily given. We have to show that there is a flat system of equations $\mathcal{E} = \langle X, A, e\rangle$, such that it holds: $B = sol(\mathcal{E})$. For the functor $\Gamma : S \longrightarrow A \times S$ we know that there is a Γ-coalgebra $\langle S, \alpha_S\rangle$, such that it holds: for every stream $s = \langle a_0, a_1, \dots\rangle \in B$, there is an infinite transition with the following dynamics:

$$s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} s_2 \xrightarrow{a_2} \dots$$

In other words: every stream taken from B can be represented as a solution set of a Γ-coalgebra $\langle S, \alpha_S\rangle$. Applying Remark 13.2.2(ii), we can deduce that there exists a flat system of streams $\langle X, A, e\rangle$, such that $B = sol(\mathcal{E})$. That suffices to show the proposition. q.e.d.

Remark 13.2.3 It is a consequence of the above facts that the union of all solution-sets of all possible flat systems of streams characterizes precisely the maximal fixed point of the operator $\Gamma : S \longrightarrow A \times S$. In other words, maximal fixed points (or final coalgebras in a more category theoretic terminology) are correlated with solution sets of systems of equations. This correspondence will be generalized in Chapter 14 where we show how to formulate the whole theory of non-well-founded sets (in which streams are only a special case) in a more general setting.

We have two alternatives when working with streams. One is to represent streams as solutions of systems as defined in Definition 13.2.2. The other alternative is to consider Γ-coalgebras and to associate streams with elements of the maximal fixed point of the operator $\Gamma : S \longrightarrow A \times S$. Why do we use different set theories? On the one hand we work in classical ZFC set theory and we consider maximal fixed points. On the other hand we work in ZFC_{afa} and consider solutions of systems of equations. Do we really need non-well-founded sets? In a certain sense, non-well-founded sets are not crucial in order to model (even infinite) streams. We can establish the whole machinery without them. In another sense, non-well-founded sets help to clarify the situation. Consider again the system of streams $\langle\{x\}, \{a\}, e\rangle$ where $e : x \longmapsto \langle a, x\rangle$. Obviously, $\langle\{x\}, \{a\}, e\rangle$ has a circular character. The solution set for this system is the stream $\langle a, a, a, ...\rangle$. Now, consider the following two deterministic labeled transition systems:

$$s_0 \xrightarrow{a} s_1 \xrightarrow{a} s_2 \xrightarrow{a} \;...$$

$$t \xleftrightarrow{a} t'$$

Obviously, the two systems are structurally equal. Moreover, they are bisimilar in the sense of Definition 13.1.6. Notice that both labeled transition systems have as output the infinite stream $\langle a, a, a, ...\rangle$. Hence, from the standpoint of the observer they are indistinguishable. Infinite streams that include a n-circle (for $n \in \mathbb{N}$) can be represented as circular structures. Important is the structural similarity, not the existence or non-existence of a bijection. Because of the fact that the (finite) circular structure is bisimilar to an infinite structure (that unfolds the circles in a certain sense), we have on the one hand a non-well-founded representation (i.e. a circular representation) and on the other hand a well-founded and infinite representation. This shows that non-well-founded sets enables us to represent certain infinite streams as finite (but circular) objects.

We can make this relation more precise by the following proposition. Intuitively, every circular transition system that has at most finitely many states can be represented as a flat system of streams with at most finitely many equations.

Proposition 13.2.4 *Assume A is an arbitrarily given set of labels. Every stream $s \in A^\omega$ can be represented as a solution of a flat system of streams with at most κ many equations, provided that the generating deterministic labeled transition system $\langle T, \longrightarrow_T, A\rangle \cong \langle T, \alpha_T\rangle$ has at most κ many states $t \in T$.*

Proof: Assume $\langle T, \alpha_T\rangle$ is a deterministic labeled transition system where $|T| \leq \kappa$, and s is a given stream. It is clear from considerations of the cardinality of states and the properties of the transition system that the transitions consist of one and the same circle (of a particular length, say α). Therefore the stream $s \in A^\omega$ can be written as

$$s = \langle a_1, a_2, \ldots, a_\alpha, a_1, a_2, ..., a_\alpha, a_1, \ldots\rangle$$

We define the system of equations $\mathcal{E} = \langle X, A, e\rangle$ with $|X| = |S|$ and $e : X \longrightarrow A \times X$, such that the following holds:

$$\begin{array}{lcl}
x_1 & \longrightarrow & \langle a_1, x_2\rangle \\
x_2 & \longrightarrow & \langle a_2, x_3\rangle \\
\vdots & \vdots & \vdots \quad \vdots \\
x_\kappa & \longrightarrow & \langle a_\kappa, x_{\kappa+1}\rangle \\
x_{\kappa+1} & \longrightarrow & \langle a_1, x_2\rangle
\end{array}$$

The solution of x_1 specifies s: $sol(x_1) = \langle a_1, a_2, \ldots, a_\alpha, a_1, a_2, \ldots \rangle = s$. This proves the claim of the proposition. q.e.d.

Remark 13.2.4 The above fact shows that relative to the functor $\Gamma : S \longrightarrow A \times S$ every infinite stream can be represented as a solution of a system of equations. Notice that we are dealing with infinite objects (streams) that cannot be defined by an ordinary (mathematical) induction. What is the connection of these considerations with circularity? Certain infinite streams are objects that are bisimilar to finite but circular objects. This is the technical representation of a philosophically important fact: whereas a circular phenomenon (compare Chapter 2) like a particular proposition, a sentence, an utterance, an abstract object like a situation etc., is a finite object, the correct understanding and interpretation of this phenomenon requires quite often an infinite regress. This infinite regress corresponds to the unfolded circle. Precisely the correct interpretation of the unfolded infinite chain makes it mathematically so difficult to handle it: we are dealing with entities that cannot be analyzed using standard mathematical tools. In particular, we cannot apply standard inductive principles in order to get an analysis of these entities.

In the next section, we will consider how streams can be manipulated, i.e. how we can define operations on infinite streams. In particular, we will examine techniques that allow us to define new objects by so-called coinductive definitions.

13.3 Operations on Infinite Objects

We have examined relationships between (flat) systems of streams, Γ-coalgebras for a functor $\Gamma : S \longrightarrow A \times S$, and labeled transition systems. We figured out that we can translate Γ-coalgebras and labeled transition systems into a system of streams. A natural question arises from the fact that we do not know how to manipulate given streams, i.e. how to define operations on streams. For example, if s_1 and s_2 are streams, how can we define a merge operation on s_1 and s_2? We will examine the possibility to introduce operations on streams in this section. The principle that will enable us to do this is the coinduction principle.

The starting point is the classical example in [BarMo96] and [Rut96] concerning labeled transition systems: the zipping (merging) of two infinite streams. This is an operation that takes two streams and creates a new stream by merging the former streams.

Example 13.3.1 Assume the functor $\Gamma : S \longrightarrow A \times S$ is given. We know from Fact 13.2.1 that A^ω is the maximal fixed point of Γ. Now consider the coalgebra $\langle A^\omega, \langle h, t \rangle \rangle$ with the following specifications:

$$h : S \longrightarrow A : s \longmapsto 1^{st}(s)$$
$$t : S \longrightarrow S : s \longmapsto 2^{nd}(s)$$

We claim that $\langle A^\omega, \langle h, t\rangle\rangle$ is a final coalgebra for Γ. Checking this fact is straightforward and left to the reader. Assume $s \in A^\omega$ and $s' \in A^\omega$ are two given streams. Assume further that a function $f : A^\omega \times A^\omega \longrightarrow A \times (A^\omega \times A^\omega)$ is given. f is defined by the following condition:

$$f(\langle s, u\rangle) = \langle h(s), \langle u, t(s)\rangle\rangle$$

The function f induces a unique operation $merge : A^\omega \times A^\omega \longrightarrow A^\omega$, such that it holds:

$$(\dagger) \quad \langle h, t\rangle(merge(a \circ v, u)) = \langle a, merge(\langle u, v\rangle)\rangle$$

The following commuting diagram clarifies the situation. The equation (†) is a consequence of this diagram.

$$\begin{tikzcd}
A^\omega \times A^\omega \arrow[r, "merge"] \arrow[d, "f"'] & A^\omega \arrow[d, "\langle h, t\rangle"] \\
A \times (A^\omega \times A^\omega) \arrow[r, "\Gamma(merge)"'] & A \times A^\omega
\end{tikzcd}$$

The unique operation $merge : A^\omega \times A^\omega \longrightarrow A^\omega$ can be described as follows. Take two streams $s = \langle a_1, a_2, ...\rangle$ and $s' = \langle b_1, b_2, ...\rangle$. Then, $merge(\langle s, s'\rangle) = \langle a_1, b_1, a_2, b_2, a_3, \ldots\rangle$. In other words, we define a new stream $merge(\langle s, s'\rangle)$ by taking the first element of s, then the first element of s', then the second element of s, and so on. We see that it is possible to define operations on infinite objects by coalgebraic means. Notice that this is not possible with inductive (hence algebraic) techniques in general.

The described procedure provides a possibility to define a function $merge$ that merges two given streams. It is important that $merge$ is not defined by classical induction, because there is no base case. The function f can be interpreted as bootstrapper of the machinery.

Notice that in our example we still consider the functor $\Gamma : S \longrightarrow A \times S$. We simply take a special coalgebra $\langle S, \alpha_S\rangle = \langle A^\omega \times A^\omega, f\rangle$ as our basis. But $A^\omega \times A^\omega$ is still a set and therefore the whole construction fits perfectly into the general account. By the finality of $\langle A^\omega, \langle h, t\rangle\rangle$ the function $merge$ must be unique and must satisfy condition (†).

We would like to give a general account for operations on streams. Additionally, we would like to compare this with the other representations of streams we saw in the above section. Concerning the first question, a generalization of the above example for arbitrary operations is needed. This generalization is called coinduction and is essentially based on the property of final coalgebras that there exists precisely one morphism from any coalgebra into the final one. The following definition makes it precise what we mean by the concept of a definition by coinduction.

Definition 13.3.1 *Assume Γ is a functor, $\langle S, \alpha_S\rangle$ is a Γ-coalgebra and $\langle \Gamma^*, \gamma\rangle$ is a final Γ-coalgebra. By the definition of finality there exists a unique homomorphism $f : S \longrightarrow \Gamma^*$, such that the following diagram commutes.*

$$\begin{tikzcd}
S \arrow[r, "\exists!\, f"] \arrow[d, "\alpha_S"'] & \Gamma^* \arrow[d, "\gamma"] \\
\Gamma(S) \arrow[r, "\Gamma(f)"'] & \Gamma(\Gamma^*)
\end{tikzcd}$$

We call $f : S \longrightarrow \Gamma^*$ defined by coinduction.

Remark 13.3.2 (i) The transition function α_S induces the whole process. It can be interpreted as the function that initiates the coinductive definition and keeps it going.

(ii) The principle of definition by coinduction as described in Definition 13.3.1 is correlated to the principle of definition by induction. Definition 13.3.1 is the dual of a definition by induction. Speaking in category theoretic terms, the induction principle can be formulated as follows. Substitute in Definition 13.3.1 the term coalgebra by the term algebra (i.e. a pair $\langle S, \alpha_S\rangle$ where $\alpha_S : \Gamma(S) \longrightarrow S$) and the term final by the term initial (i.e. consider an algebra $\langle \Gamma_*, \gamma\rangle$ instead of a coalgebra $\langle \Gamma^*, \gamma\rangle$), then there exists a unique homomorphism $f' : \Gamma_* \longrightarrow S$, such that the following diagram commutes:

$$\begin{tikzcd}
\Gamma(S) \arrow[d, "\alpha_S"'] & \Gamma(\Gamma_*) \arrow[l, "\Gamma(f')"'] \arrow[d, "\gamma"] \\
S & \Gamma_* \arrow[l, "\exists!\, f'"]
\end{tikzcd}$$

Whereas the presented abstract form of the induction principle is used overall in mathematics, one of the most prominent areas where this principle is essentially the basis is recursion theory. In recursion theory, Γ_* is the minimal fixed point of the ordinary successor function, i.e. the set of natural numbers $\mathbb{N}$ (roughly speaking) and the induction principle enables us to define functions from the natural numbers into other sets. Examples are elementary arithmetical operations. Consider the functor $\Gamma : \langle S, S\rangle \longrightarrow 1 + S$. Assume a Γ-algebra $\langle S, \alpha_S\rangle$ is given where the dynamics α_S is defined by:

$$\begin{aligned}&\alpha_S(\langle *, *\rangle) = 0\\ &\alpha_S(\langle x, *\rangle) = x\\ &\alpha_S(\langle x, suc(y)\rangle) = suc(\alpha_S(\langle x, y\rangle))\end{aligned}$$

This is precisely the classical recursive definition of addition. Notice that the Γ-algebra $\langle \mathbb{N}, 0, suc\rangle$ is initial for that functor. Notice further that given two Γ-algebras $\langle S, \alpha_S\rangle$ and $\langle T, \alpha_T\rangle$ a homomorphism $f : S \longrightarrow T$ is a function, such that the following well-known properties for addition do hold:

$$\begin{aligned}&f(0_S) = 0_T\\ &f(\langle x, y\rangle_S) = \langle f(x), f(y)\rangle_T\end{aligned}$$

This shows how the account of Γ-coalgebras can be seen as the dual theory of the theory of Γ-algebras.

We mentioned a second concern above, namely the relation between the coalgebraic account for deterministic labeled transition systems and the representation of labeled transition systems as solutions of systems of streams. This is straightforward to show using Remark 13.2.2(ii). We mirror everything we did already there. If we consider two streams $s \in A^\omega$ and $t \in A^\omega$ and a function $\alpha : A^\omega \times A^\omega \longrightarrow A \times (A^\omega \times A^\omega)$ is given defined by the following condition:[3]

$$\alpha(\langle s, u\rangle) = \langle h(s), \langle u, t(s)\rangle\rangle$$

then we are able to represent the unique homomorphism $merge : A^\omega \times A^\omega \longrightarrow A^\omega$ by the following system of equations. Take a set of urelements $X \subseteq \mathcal{U}$, such that there is a bijective mapping $g : X \longrightarrow A^\omega \times A^\omega$. We define $e : X \longrightarrow \wp(X \cup A)$ according to the following condition:

$e(x) = \langle a, x'\rangle$ iff the following three conditions (i) - (iii) hold:

(i) $\alpha(\langle s, u\rangle) = \langle h(s), \langle u, t(s)\rangle\rangle$
(ii) $h(s) = a$
(iii) $\alpha(\langle u, t(s)\rangle) = x'$

[3] Notice that h and t are the functions defined in Example 13.3.1

Although the above argumentation is a quite specific example of merging two streams it is clear that this procedure works for many kinds of operations on streams in a quite general way. We simply translate what we do in the theory of coalgebras into the theory of systems of streams.

In this section, we considered the internal connection between the theory of non-well-founded sets and the theory of Γ-coalgebras. Important is the fact that a circular object modeled as a solution of a system of (set theoretical) equations is represented as the unfolded version of this circular object. The circular object is transformed to an infinite one in this process.

Our presentation in this section was quite specific with respect to the functor we considered. In the next section, we shall consider the relation between systems of equations and final coalgebras more closely.

13.4 Final Coalgebras and Systems of Equations

Deterministic labeled transition systems are automata that were extensively studied, because of the large variety of applications in logic, computer science, and linguistics. Recently, there was the endeavor to find coalgebraic representations of a variety of other automata. One can give a coalgebraic representation of non-deterministic labeled transition systems, (deterministic or non-deterministic) finite state automata, transducers, or tree automata. The difficulties behind these approaches is quite often the same. One has to find the correct functor Γ, such that the induced Γ-coalgebras are in one-to-one correspondence with the intended type of automata. In order to make this more precise, we give an example.

Example 13.4.1 Consider a finite state automaton $\mathbf{S} = \langle S, \Sigma, a_0, F, \delta \rangle$ where S is a finite set of states, Σ is a finite alphabet, a_0 is the state in which the automaton starts, F is a collection of final states, and $\delta : S \times \Sigma^+ \longrightarrow S$ is the transition function. Consider the functor $\Gamma : \mathcal{SET} \longrightarrow \mathcal{SET} : S \longrightarrow \mathbf{1} + S^\Sigma$. The induced Γ-coalgebras $\langle S, \alpha_S \rangle$ can be easily specified as follows:

$$(^*) \qquad s \xrightarrow{a} s' \iff \alpha_S(s)(a) = s'$$

We claim that every finite automaton can be represented as a Γ-coalgebra and vice versa. Notice first that every finite automaton with an initial state can be represented equivalently as an automaton with more than one initial state. This is a well-known result of the theory of automata.[4] Therefore, the specification of the initial state a_0 is superfluous. Concerning the collection of final states, it is clear that this information is internally captured in the transition function α_S. That is the reason why we can model an automaton without specifying explicitly the final states, because we assume that this information is internally given by the transition function. Consider the definition of the dynamics of the coalgebra. If the automaton reads the symbol

[4] Compare for example [HoUl79].

a while being in state s, the automaton performs the transition to state s'. We model the behavior of the automaton precisely with the above condition (*).[5]

The above considerations concerning finite state automata make clear that every representation depends on the choice of the correct functor Γ. Each functor Γ that induces a final coalgebra can be interpreted as a construction method to define types of automata. The problem is to guarantee that the functor induces a final coalgebra. We want to consider this problem more closely keeping in mind that we want to find a more general account of the correspondence between final coalgebras and systems of equations.

We examine conditions that guarantee that certain functors do have final coalgebras. Notice that not all functors guarantee the existence of final objects. For example, we know that the category of all Γ-coalgebras where Γ is given by $\Gamma : S \longrightarrow \wp(S)$ has no final coalgebra, because the existence contradicts Cantor's theorem.

Intuitively, such categories do have final coalgebras that are in a certain sense bound. For example, functors that are composed of polynomial operations like $+$ (sum), $\times$ (product), I (identity),$(-)^A$ (function space) are bound and do have final coalgebras. We saw already certain examples where some of these operations were involved. Another example for a functor that generates final coalgebras is $\wp_{fin}$ (compare Theorem 13.1.10). There is a general fact that guarantees that certain coalgebras do have final ones. Although there is no sufficient and necessary condition known that guarantees the existence of final coalgebras, we can give at least sufficient conditions. Before we are able to do this first, we need two definition.

Definition 13.4.1 *Assume* $\Gamma : \mathcal{SET} \longrightarrow \mathcal{SET}$ *is a functor. Assume* $\mathbf{1}$ *is the set with one element. Let* $!$ *be the unique function with the property* $! : \Gamma(\mathbf{1}) \longrightarrow \mathbf{1}$*. For* $\Gamma^{n+1} = \Gamma \circ \Gamma^n$ *we define the inverse limit of the sequence*

$$\mathbf{1} \xleftarrow{!} \Gamma(\mathbf{1}) \xleftarrow{\Gamma(!)} \Gamma^2(\mathbf{1}) \xleftarrow{\Gamma^2(!)} \ldots$$

as the following set P*:*

$$P = \{\langle x_0, x_1, x_2, \ldots \rangle \mid \forall n \in \omega : x_n \in \Gamma^n(\mathbf{1})\} \;\wedge\; \Gamma^n(!)(x_{n+1}) = x_n$$

Definition 13.4.2 *Assume* $\Gamma : \mathcal{SET} \longrightarrow \mathcal{SET}$ *is a given functor and* P *is its inverse limit. We call* Γ $\omega^{\mathcal{OP}}$*- continuous, if* $\Gamma(P)$ *is again a limit of the sequence*

$$\mathbf{1} \xleftarrow{!} \Gamma(\mathbf{1}) \xleftarrow{\Gamma(!)} \Gamma^2(\mathbf{1}) \xleftarrow{\Gamma^2(!)} \ldots$$

[5] More information concerning the representation of a variety of automata using coalgebras can be found in [Ku∞a].

Using these concepts it is possible to formulate a construction for final coalgebras. Assume P is an inverse limit, and additionally, $\Gamma(P)$ is again a limit of the same sequence. Then, there exists a bijection from P to $\Gamma(P)$. Hence, the inverse function $\pi : \Gamma(P) \longrightarrow P$ of that bijection provides the dynamics of the final coalgebra $\langle P, \pi\rangle$.

In the following fact, we state sufficient conditions for a functor Γ to have final coalgebras:

Fact 13.4.3 *Assume the category $\mathcal{SET}$ is given. All functors $\Gamma : \mathcal{SET} \longrightarrow \mathcal{SET}$ that are build using only operations of the set $\{+, \times, I, (-)^A, \wp_{fin}\}$ do have final coalgebras.*

Proof: The idea is to check that the functors build from the operations $\{+, \times, I, (-)^A\}$ are $\omega^{\mathcal{OP}}$-continuous. This can easily be achieved. Consider the Γ-coalgebra $\langle P, \pi\rangle$ where $\pi : P \longrightarrow \Gamma(P)$ is defined as the inverse function of a bijective function $f : \Gamma(P) \longrightarrow P$. f exists because Γ is $\omega^{\mathcal{OP}}$-continuous. It remains to show that $\langle P, \pi\rangle$ is a final coalgebra. This can be easily shown.
q.e.d.

All functors we are considering in this section are build from polynomial operations or the finite power set operation. We would like to represent an arbitrary Γ-coalgebra by a system of set theoretical equations. Additionally, we want to represent the maximal fixed point of the operator as a system of equations. First, we give some simple examples.

Example 13.4.2 (i) Assume $\Gamma : \mathcal{SET} \longrightarrow \mathcal{SET}$ is specified as follows: $\Gamma : S \longrightarrow A \times S$. As we saw in Section 13.2, the Γ-coalgebra $\langle A^\omega, \langle h, t\rangle\rangle$ is final. Furthermore, we saw in that section how we can interpret each stream as a solution of a system of equations.

(ii) Assume $\Gamma : \mathcal{SET} \longrightarrow \mathcal{SET}$ is specified as follows: $\Gamma : S \longrightarrow \mathbf{1} + (\mathbf{1} \times S)$. Let $\mathbb{N}_\infty$ be defined as follows: $\mathbb{N}_\infty = \mathbb{N} \cup \{\omega\} = \{0, 1, 2, ...\} \cup \{\omega\}$. Notice that the Γ-coalgebra $\langle \mathbb{N}_\infty, pred\rangle$ is a final coalgebra for Γ provided that the dynamics $pred$ is defined as follows:

$$pred = \begin{cases} * & : \; n = 0 \\ n-1 & : \; 0 < n \wedge n \neq \omega \\ \omega & : \; n = \omega \end{cases}$$

Every natural number n makes a transition to its predecessor (and terminates after n steps), whereas ω never terminates and is mapped to ω again. Clearly, for every Γ-coalgebra $\langle S, \alpha_S\rangle$ there exists precisely one arrow to $\langle \mathbb{N}_\infty, pred\rangle$. Trivially, every subset $X \subseteq \mathbb{N}_\infty$ can be represented as a system of set theoretical equations. For example, $\mathbb{N}_\infty$ itself is the solution of the system $\langle X, A, e\rangle$ where

$X = \{x_0, x_1, x_2, ...\} \cup \{y\}$, $A = \emptyset$ and $e : X \longrightarrow \wp(X)$ is defined as follows:

$$\begin{aligned}
&e(x_0) = \{\emptyset\}\\
&e(x_1) = \{x_0\}\\
&e(x_2) = \{x_1\}\\
&\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots\\
&e(y) = \{x_0, x_1, x_2, ...\}
\end{aligned}$$

Clearly, the solution of $\langle X, A, e\rangle$ gives us precisely the set $\mathbb{N}_\infty$.

We want to give a general account for modeling subsets of final coalgebras as solutions of systems of equations using the set theoretical anti-foundation axiom. We can argue in a quite general setting. We will consider the following functor $\Gamma : \mathcal{SET} \longrightarrow \mathcal{SET}$ specified as follows:

$$(^{**}) \qquad \Gamma(S) = (A_1 + B_1 \times S)^{C_1} \times (A_2 + B_2 \times S)^{C_2} \times \cdots \times (A_n + B_n \times S)^{C_n}$$

This functor includes the operation $+, \times, (-)^X$ and $\mathbf{1}$ (termination of the automaton). We did not use the finite power set operation, but clearly this operation could be added without any further complications. The next fact points out how to model subsets of the final coalgebra of such a functor as a system of set theoretical equations. That comes down to a straightforward generalization of Proposition 13.2.3.

Fact 13.4.4 *Assume Γ is defined as in (**). Assume further that $\langle P, \pi\rangle$ is the final coalgebra for Γ. Then it holds: For every subset $P' \subseteq P$ there exists a system of equations $\mathcal{E}$, such that the solution-set of $\mathcal{E}$ is equal to P'.*

Proof: The collection of all solution-sets of systems of equations equals the whole non-well-founded universe. In particular, one can reach every set in the class of all sets. q.e.d.

There is a deeper connection between systems of equations and maximal fixed points. The following proposition gives a characterization of certain types of systems of equations.

Proposition 13.4.5 *Consider the following system of equations $\mathcal{E}$ specified by the equation:*

$$x = (A_1 + B_1 \times x)^{C_1} \times (A_2 + B_2 \times x)^{C_2} \times \cdots \times (A_n + B_n \times x)^{C_n}$$

Then it holds: the solution $s(x)$ specifies precisely the collection of sets specified by the maximal fixed point Γ^* for Γ given by (**).

Proof: We need to show that it holds: $s(x) = \Gamma^*$. It suffices to show the claim of a simplified version of Γ. Assume $n = 1$ and $C_1 = 1$. Then, we need to show that it holds:

$$s(x) = \langle A_1 + B_1 \times s(x) \rangle \; = \; \Gamma^*$$

where $\Gamma : S \longrightarrow A_1 + B_1 \times S$. Assume $y \in s(x)$. Then, y is of the form:

$$y = \langle \{ \langle 0, a_\alpha \rangle, \langle 1, \langle b_\beta, y \rangle \rangle \} \rangle$$

Then, clearly it holds: $y \in \Gamma^*$. This shows one direction of the proposition.

For the other direction, assume $y \in \Gamma^*$. Then, we get:

$$y = \langle \{ \langle 0, a_\alpha \rangle, \langle 1, \langle b_\beta, s' \rangle \rangle \} \rangle$$

Because s' has again the correct form, it follows immediately: $y \in s(x)$. This suffices to show the proposition. q.e.d.

We finish this chapter with these remarks. In a certain sense, the considerations in this chapter are not very surprising, because it is clear that everything that is a set in the non-well-founded universe can be constructed using set theoretic equations. This holds, because the collection of all set theoretic equations precisely defines the non-well-founded universe.

In the following chapter, we will use the coalgebraic approach quite extensively in order to give a general account for the theory of non-well-founded sets. In a certain sense, the next chapter is a generalization of the considerations developed so far. Clearly, becoming more general means at the same time becoming more abstract.

13.5 History

The concept of coalgebras and streams are well-known in computer science. The approach developed here is quite close to the explanations given in [Rut96, BarMo96], and [Ja95b] where several of the ideas are taken from. Further references and information can be found in [Ja95a], [Ja96a], and [Ja96b]. Concerning coalgebras and their applications in a wider context one find a lot of material in [CMCS98]. A good source concerning the existence of final systems can be found in [Rut96]. The first textbook that uses coalgebras extensively, in order to give an account for theoretical computer science and, in particular, to give an account for the semantics of programming languages was [ManArb86]. The relationship between labeled transition systems, Γ-coalgebras, and the solution set of systems of equations goes back to [BarMo96]. An examination and comparison of infinite but well-founded structures and finite but circular structures as well as the different forms of representing operations on them is, as far as the author knows, not discussed in the literature. Section 13.4 of this chapter generalizes the account given in [BarMo96] where systems of equations are related to final coalgebras. A good textbook for labeled transition systems from the perspective of dynamic logic is [Be96].

Chapter 14

Non-Well-Founded Sets and Coalgebras

In this chapter, we will generalize the considerations developed so far. We will begin this chapter with an examination of a generalization of the corecursive substitution principle as stated in Theorem 11.6.2. By applying standard techniques of category theory and the theory of coalgebras it is possible to prove the existence and the uniqueness of the corecursive substitution principle quite easily in a more general setting. The possibility to flatten a set theoretic system of equations is based on the finality property of final coalgebras and the correct choice of the categroy theoretic construction, namely a construction that is mainly based on coproducts. As a further step in the development we will consider the corecursion theorem in its parameterized version. This chapter is the basis for the coalgebraic treatment of situation theory in Section 15.2. Our presentation is quite close to the ideas formulated in [Mo∞b] although we emphasize strongly the set theoretical motivation. Moreover, the usage of urelements is a difference between our presentation and [Mo∞b].

14.1 Flattening

We saw in Theorem 11.6.2 that the corecursive substitution operation $sub(s, x)$ for a substitution s defined on urelements and an element $x \in \mathcal{U} \cup \mathcal{V}_{afa}[\mathcal{U}]$ exists and is uniquely defined. The proof of this theorem is rather complicated and not straightforward. We would like to develop an alternative account in a more abstract setting. A direct translation of the substitution operation to the coalgebraic case is not possible. Fortunately, we can use another construction. The non-well-founded universe (constructed relatively to a collection of urelements) can play the role of a final coalgebra. Because of the finality of the universe we can establish the existence and uniqueness of a morphism into this final coalgebra. Hence, the result of a unique corecursive substitution principle follows for nothing (under the assumption that the anti-foundation axiom holds).

The following fact is the general case of flattening a system for an arbitrary endofuncor $\Gamma : \mathcal{SET} \longrightarrow \mathcal{SET}$ where Γ induces a final coalgebra.

Fact 14.1.1 *Assume $\langle S, \alpha_S \rangle$ and $\langle T, \alpha_T \rangle$ are Γ-coalgebras for a given functor $\Gamma : \mathcal{SET} \longrightarrow \mathcal{SET}$, such that Γ induces a final coalgebra.*[1] *Assume further that $\oplus$ is a coproduct operation and that $\langle P, \pi \rangle$ is a final coalgebra. If $f : T \longrightarrow \Gamma(S \oplus T)$ is an arbitrary function, then there are unique functions $g : T \longrightarrow P$ and $\gamma = \langle g', g \rangle : S \oplus T \longrightarrow P$, such that it holds:*

$$\alpha_P \circ g \ = \ \Gamma(\langle g', g \rangle) \circ f$$

Proof: Notice that the functions $g : T \longrightarrow P$ and $\gamma : S \oplus T \longrightarrow P$ exist by finality of $\langle P, \pi \rangle$. We need to show that it holds: $\alpha_P \circ g = \Gamma(\gamma) \circ f$. The following diagram makes the situation clear:

$$\begin{tikzcd}
 & \Gamma(P) & \\
 & P \arrow[u, "\alpha_P"] & \\
S \arrow[ur, "g'"] \arrow[r, "in_l"'] \arrow[d, "\alpha_S"'] & S \oplus T \arrow[u, "{\langle g', g \rangle}"'] \arrow[d, "\alpha_{S \oplus T}"'] & T \arrow[ul, "g"'] \arrow[l, "in_r"] \arrow[dl, "f"] \arrow[d, "\alpha_T"] \\
\Gamma(S) \arrow[r, "\Gamma(in_l)"'] \arrow[uuur, bend left, "\Gamma(g')"] & \Gamma(S \oplus T) & \Gamma(T) \arrow[l, "\Gamma(in_r)"] \arrow[uuul, bend right, "\Gamma(g)"']
\end{tikzcd}$$

Because $\langle P, \pi \rangle$ is final we know that $\gamma = \langle g', g \rangle : S \oplus T \longrightarrow P$ is a uniquely defined arrow. Furthermore, by the definition of a coproduct it holds: $in_l \circ \gamma = g'$. Concerning the properties of f, we have the following equalities:

$$\begin{aligned}
\Gamma(\gamma) \circ f \ &= \ \Gamma(\gamma) \circ (\alpha_{S \oplus T} \circ in_r) \\
&= \ \alpha_P \circ (\gamma \circ in_r) \\
&= \ \alpha_P \circ g
\end{aligned}$$

[1] For example, in the case Γ is built by the operations $\{+, \times, I, (-)^A, \wp_{fin}\}$ we know by Fact 13.4.3 that there exists a final coalgebra for Γ.

The first equality is clear. The second follows by finality of $\langle P, \pi \rangle$, and the third equality follows by the coproduct definition. The properties that the the arrows g and γ are unique are again a consequence of the finality of $\langle P, \pi \rangle$. That suffices to show the claim of the fact. q.e.d.

We add some remarks concerning Fact 14.1.1.

Remark 14.1.1 (i) Fact 14.1.1 is called flattening lemma in [Mo∞b]. The justification is that systems of equations with a deep structure can be transformed into flat systems using Fact 14.1.1. The idea is similar to the examples we saw in Section 11.6 where we explicitly transformed systems with a deep structure into flat systems. Example 14.1.2 will make this connection precise.

(ii) In the proof of Fact 14.1.1, we did not draw the arrow $\Gamma(\gamma) : \Gamma(S \oplus T) \longrightarrow \Gamma(P)$, although this arrow is necessary (and important) for the construction and the existence of this arrow is clear from the theory of coalgebras. But the readability of the diagram would suffer if we included this arrow. That is the reason why we skipped $\Gamma\gamma$ in the diagram.

(iii) The intuition of the above fact can be formulated as follows: if a system of set theoretical equations is given by a function $f : T \longrightarrow \Gamma(S \oplus T)$ and this system is not flat, then Fact 14.1.1 tells us that it is possible to transform it into a system induced by $g : T \longrightarrow \mathcal{V}_{afa}$, such that this system is in fact a flat system. Notice that we take as final coalgebra the class $\mathcal{V}_{afa}$ of the non-well-founded universe together with the identity mapping $id : \mathcal{V}_{afa} \longrightarrow \mathcal{V}_{afa}$. Notice that this intuition is not literally captured by Fact 14.1.1: We cannot work in $\mathcal{SET}$, because $\mathcal{V}_{afa}$ is not contained in $\mathcal{SET}$. Furthermore, we want to include urelements. Section 14.2 will discuss necessary modifications in order to guarantee that nothing goes wrong when working in $\mathcal{CLASS}$.

(iv) In Definition 12.2.4, we introduced coproducts where the unique function $\gamma : S \oplus T \longrightarrow P$ was denoted by the pair $\langle g', g \rangle$. We used γ as a shorthand for the pair $\langle g', g \rangle$, because of the simpler notational form.

In order to see the connection between Fact 14.1.1 and the theory of non-well-founded sets more clearly (and in order to make the intuition of Remark 14.1.1 (iii) more precise), we consider the following example.

Example 14.1.2 Assume we work in the category $\mathcal{CLASS}$ of all classes.[2] The reason for the need of $\mathcal{CLASS}$ instead of the category of all sets is the fact that we want to guarantee the existence of a final coalgebra for the power set functor $\wp$. In Proposition 14.2.1, we will show that the coalgebra $\langle \mathcal{V}_{afa}, id \rangle$ is in fact a final coalgebra for the functor $\wp$. Assume further that $S = \mathcal{V}_{afa}$,

[2] For further information concerning $\mathcal{CLASS}$ the reader is referred to the remarks below, especially to Remark 14.2.1 and the explanations in Section 14.2.

that $g' = \alpha_{\mathcal{V}_{afa}} = id_{\mathcal{V}_{afa}}$ and that the coproduct operation $\oplus$ is ordinary set theoretic disjoint union. Now we consider the following system of set theoretical equations, relative to a given set $T = \{x, y, z, \Omega, 1\}$.

$$\begin{aligned} f(x) &= \{\langle 0, 1\rangle, \langle 1, y\rangle\} \\ f(y) &= \{\langle 0, \Omega\rangle, \langle 1, z\rangle, \langle 2, y\rangle\} \\ f(z) &= \{\langle 0, x\rangle\} \end{aligned}$$

First, notice that it is impossible to apply the anti-foundation axiom in its ordinary form to the above system, because the considered equations are not flat. In Chapter 11, we examined techniques that enabled us to reduce these systems to flat systems. Now the existence (and uniqueness) of the mapping $g : T \longrightarrow \mathcal{V}_{afa}$ of Fact 14.1.1 gives us a flat system of the following form.

$$\begin{aligned} g(x) &= \{1, y\} \\ g(y) &= \{\Omega, z, y\} \\ g(z) &= \{x\} \end{aligned}$$

It is important to notice that the following equalities do hold (similar to the claim in Fact 14.1.1):

$$g = id \circ g = \Gamma(\gamma) \circ f = \Gamma(\langle id, g\rangle) \circ f$$

This is precisely what we need in order to reduce an arbitrary system of equations to a flat system of equations. In this respect, the coalgebraic account is a generalization of the techniques we examined in Chapter 11.

Remark 14.1.3 Notice that we generalized the picture of Fact 14.1.1 slightly in Example 14.1.2. We considered a coproduct operation of three coalgebras, not only of two coalgebras. According to our development of operations in categories, in particular, of coproducts in categories in Section 12.2 there is a natural generalization to an arbitrary number of coproducts. This does not change anything of the theory.

Because we worked in our example above in $\mathcal{CLASS}$ (and in Fact 14.1.1 in $\mathcal{SET}$), it is necessary to add some remarks concerning the existence of final coalgebras in $\mathcal{CLASS}$. The following section summarizes the most important facts concerning this point.

14.2 Final Coalgebras

For the further development we need some facts concerning final coalgebras and maximal fixed points. We provide some background information concerning appropriate coalegebras in this section. First, we make some comments concerning the category $\mathcal{CLASS}$.

Remark 14.2.1 (i) An important difference between the above Example 14.1.2 and the situation in Fact 14.1.1 is the choice of the category $\mathcal{CLASS}$ instead of the category $\mathcal{SET}$ on which the endofunctor $\wp$ operates. Clearly, the theory of coalgebras we developed so far was mainly formulated as a theory defined in the category $\mathcal{SET}$. As we saw in Section 13.1, there is no final coalgebra for $\wp$ in $\mathcal{SET}$. And without the existence of the final coalgebra $\langle P, \pi \rangle$ there is no chance to prove a claim like Fact 14.1.1. That is the reason why we chose the category of all classes $\mathcal{CLASS}$.

(ii) We assume that the category $\mathcal{CLASS}$ is specified as the collection of all objects, such that elements of these objects are sets. Hence, $\mathcal{CLASS}$ contains all sets and objects of the form $\{a_1, a_2, \dots\}$ where each a_α is a set. Notice that ORD is in $\mathcal{CLASS}$, because every element in ORD is a set, but $\{ORD\} \notin \mathcal{CLASS}$, because ORD itself is not a set. Similarly, $\mathcal{V}_{afa} \in \mathcal{CLASS}$, because every element of $\mathcal{V}_{afa}$ is a set, but it holds: $\{\mathcal{V}_{afa} \notin \mathcal{CLASS}\}$. Concerning morphisms we assume that every morphism $f : a \longrightarrow b$ is set continuous. Formally, this means that the following condition holds:

$$f(a) \;=\; \bigcup \{f(a') \;\mid\; a' \subseteq a \text{ such that } a' \text{ is a set}\}$$

A natural generalization is to incorporate urelements. It is straightforward to consider $\mathcal{CLASS}[X]$ instead of $\mathcal{CLASS}$. In Fact 14.2.3 we will prove the basic property that this object is final for a certain operator. Notice that $\mathcal{CLASS}[X]$ is the collection of all classes as objects that have sets or urelements as elements and are generated by $X \subseteq \mathcal{U}$. Morphisms are similarly defined as above.

The following proposition guarantees that moving from $\mathcal{SET}$ to $\mathcal{CLASS}$ does not cause any problems in our context here.[3] We will see below that some slight modifications are necessary if we incorporate urelements.

Proposition 14.2.1 *Assume we are working in the category $\mathcal{CLASS}$ specified as the collection of objects that contains proper sets as elements. Consider the power set functor $\wp : \mathcal{CLASS} \longrightarrow \mathcal{CLASS}$. If the anti-foundation axiom holds, then $\mathcal{V}_{afa}$ is a maximal fixed point of the functor $\wp$. Furthermore, the pair $\langle \mathcal{V}_{afa}, id \rangle$ is a final coalgebra for $\wp$.*

Proof: We show that $\mathcal{V}_{afa}$ is a maximal fixed point for $\wp$, i.e. we show the following equality: $\mathcal{V}_{afa} = \wp(\mathcal{V}_{afa})$. Assume $a \in \mathcal{V}_{afa}$. Because of the definition of $\wp$ it holds: $a \in \wp(a)$ and therefore $a \subseteq \mathcal{V}_{afa}$. Hence, $a \in \bigcup_{b \subseteq \mathcal{V}_{afa}} \wp(b)$ and therefore $a \in \wp(\mathcal{V}_{afa})$ because of the definition of $\wp$. For the other direction assume $a \in \wp(\mathcal{V}_{afa})$. Then, there exists a set $b \subseteq \mathcal{V}_{afa}$, such that $a \in \wp(b)$. Hence, $a \subseteq b$ for a set b of pure sets. Conclude: $a \in \mathcal{V}_{afa}$. Together we have: $\mathcal{V}_{afa} = \wp(\mathcal{V}_{afa})$.

[3] Clearly, it is impossible to transform everything from $\mathcal{SET}$ to $\mathcal{CLASS}$.

It remains to show that $\langle \mathcal{V}_{afa}, id \rangle$ is a final coalgebra for $\wp$. Using the anti-foundation axiom this is easy to show. In the following diagram $e : s \longrightarrow \wp(s)$ can be considered as a generalized system of equations ($A = \emptyset$). Now, the anti-foundation axiom guarantees that this system has a unique solution sol using Proposition 11.3.1.

$$\begin{tikzcd}
s \arrow[r, "sol"] \arrow[d, "e"'] & sol(s) \arrow[d, "id"] \\
\wp(s) \arrow[r, "sol"'] & sol(s)
\end{tikzcd}$$

According to the diagram the existence and the uniqueness of the mapping $sol : s \longrightarrow sol(s)$ implies that $\mathcal{V}_{afa}$ is final, because for every set s there is precisely one morphism into $\mathcal{V}_{afa}$, namely the solution of the generalized system. Because the collection of all solutions determines precisely the universe $\mathcal{V}_{afa}$, the claim follows. q.e.d.

The following remarks add some further information concerning certain extensions of Proposition 14.2.1.

Remark 14.2.2 (i) The situation changes if we take urelements into account. The argument in proposition 14.2.1 does only hold for the collection (class) of proper sets considered as objects of the category $\mathcal{CLASS}$. It does not hold, for example, for the class $\mathcal{V}_{afa}[\mathcal{U}]$:

$$\mathcal{V}_{afa}[\mathcal{U}] \;=\; \wp(\mathcal{V}_{afa}[\mathcal{U}])$$

In order to see this, assume $x \in \mathcal{U}$ is an urelement. Then, $\{x\} \in \mathcal{V}_{afa}[\mathcal{U}]$, but we do not have that $\{x\} \subseteq \mathcal{V}_{afa}[\mathcal{U}]$, because x is not a set. Hence, it holds $\mathcal{V}_{afa}[\mathcal{U}] \not\subseteq \wp(\mathcal{V}_{afa}[\mathcal{U}])$. Therefore, it is important that there exists a unique solution of generalized systems of equations in order to get the desired result. Furthermore, it is clear that some modifications are needed if we want to develop a version of Proposition 14.2.1 for a universe that includes urelements. We will discuss these modifications below.

(ii) It is helpful to mention an important difference between the maximal fixed points of monotone operators we considered in Section 7.4 (in the context of coinductive definitions) and the maximal fixed points we consider in this chapter. In Section 7.4, we worked in ordinary ZFC set theory without allowing reflexive sets. In the context of the theory of coalgebras, we work in ZFC_{afa}. Relative to the underlying set theory differences occur concerning the features of minimal and maximal fixed points of the operator $\wp$. Whereas relative to ZFC the minimal and the maximal fixed point of $\wp$ is the same collection, namely the universe $\mathcal{V}$ of well-founded sets, in the case of ZFC_{afa} the situation changes. The minimal fixed point remains $\mathcal{V}$, but the maximal fixed point turns out to be different from $\mathcal{V}$, namely the collection $\mathcal{V}_{afa}$ of all well-founded and non-well-founded sets. That makes clear that we need to keep distinct things distinct.

Although we can reduce arbitrary systems of equations to flat systems, we still do not know whether an appropriate notion of corecursive substitution can be implemented into the coalgebraic framework. In order to establish this concept, we work again in the category $\mathcal{CLASS}$ in which every object of that category has proper sets as elements. The coproduct operation we apply is disjoint union, and the functor Γ is taken to be a modified power set operation $\wp_X$ for a collection of urelements $X \subseteq \mathcal{U}$. The operator $\wp_X$ is defined as follows:

$$\wp_X : a \longmapsto \wp(a) \cup \mathit{support}(a)$$

Intuitively, $\wp_X$ is the ordinary power set operations on collections that include additionally flat sets where the elements of such sets are urelements that are somehow involved in building the collection a. Clearly, the modified power set operation $\wp_X$ is intended to avoid the problems sketched in Remark 14.2.2(i) when urelements are included.

Because the formulation of the theory of hypersets we developed in Chapter 11 uses urelements as basic entities to build the intended universe $\mathcal{V}_{afa}[\mathcal{U}]$, we also need to introduce urelements. Therefore, we assume that there is a (proper) class of urelements $\mathcal{U}$. We specify the maximal fixed point of the modified power set operation $\wp_X$ defined above in the following definition.

Definition 14.2.2 *Assume $\Gamma : \mathcal{V}_{afa}[X] \longrightarrow \mathcal{V}_{afa}[X]$ is the modified power set functor $\wp_X$ for a collection of urelements $X \subseteq \mathcal{U}$ as sepcified above. We denote by X^* the maximal fixed point of $\wp_X$ provided this fixed point exists.*

At this point of the development it is not clear whether the maximal fixed point X^* exists for arbitrary collections $X \subseteq \mathcal{U}$. In order to justify the above definition we state Fact 14.2.3 where this existence is guaranteed. But Fact 14.2.3 proves even more: it shows that the maximal fixed point together with the identity morphism is also a final coalgebra the modified power set functor $\wp_X$.

Fact 14.2.3 *(i) The collection $\mathcal{V}_{afa}[X]$ is a maximal fixed point X^* of $\wp_X$ provided the anti-foundation axiom holds.*
(ii) For every collection $X \subseteq \mathcal{U}$, the pair $\langle \mathcal{V}_{afa}[X], id_{\mathcal{V}_{afa}[X]} \rangle$ is a final $\wp_X$-coalgebra provided the anti-foundation axiom holds.

Proof: (i) The claim is a generalization of Proposition 14.2.1 making a statement not only about pure sets but also about sets that contain urelements. For pure sets the argument of Proposition 14.2.1 can directly be applied. We need to consider the problematic case in which urelements are involved. The crucial case is $\{x\} \in \mathcal{V}_{afa}[X]$ (compare Remark 14.2.2(i)). Clearly we have $\{x\} \in \wp_X(\mathcal{V}_{afa}[X])$. Hence, it holds: $\mathcal{V}_{afa}[X] \subseteq \wp_X(\mathcal{V}_{afa}[X])$. The other direction is clear, because every $a \in \wp_X(\mathcal{V}_{afa}[X])$ is obviously a set in $\mathcal{V}_{afa}[X]$. Conclude: $\mathcal{V}_{afa}[X]$ is in fact a fixed point. Because $\mathcal{V}_{afa}[X]$ is the whole non-well-founded universe there are no larger set theoretic collections, hence $\mathcal{V}_{afa}[X]$ is also maximal.

(ii) It remains to show that $\langle \mathcal{V}_{afa}[X], id_{\mathcal{V}_{afa}[X]} \rangle$ is a final coalgebra for $\wp_X$. In other words, we need to show that for every class (with sets as elements) there is precisely one morphism into $\langle \mathcal{V}_{afa}[X], id_{\mathcal{V}_{afa}[X]} \rangle$. Here, we can apply the anti-foundation axiom, and the argument is similar to the one of Proposition 14.2.1 above. q.e.d.

We add some remarks concerning the above fact, in order to make some ideas more transparent.

Remark 14.2.3 (i) The motivation for constructing the maximal fixed point $X^* = \mathcal{V}_{afa}[X]$ is the necessity to introduce a coding scheme. This is quite similar to the usage of variables in the case of the standard representation of the theory of hypersets as was developed in Chapter 11. Variables can be interpreted as a code of a set a. The solution for a system of equations can be interpreted as an assignment of a set of the universe to every code.[4]

(ii) Notice that Fact 14.2.3 is not true if one works with the ordinary power set functor $\wp$. The problem is the status of the urelements as non-sets. The slight modification of the power set functor solves this problem. Although the usage of urelements makes the development of the theory more complicated, urelements are quite important in applications as we will see in Chapter 15.

With these remarks we can formulate the specialized form of Fact 14.1.1 that is appropriate for our further development.

Fact 14.2.4 *Assume $\langle X, \alpha_X \rangle$ and $\langle Y, \alpha_Y \rangle$ are $\wp_X$- and $\wp_Y$ coalgebras for $\wp_Z : \mathcal{CLASS}[\mathcal{U}] \longrightarrow \mathcal{CLASS}[\mathcal{U}]$. If $f : X \longrightarrow \wp_X(Y \oplus X)$, then there exists a unique morphism $g : X \longrightarrow \mathcal{V}_{afa}[X]$ and $\gamma : Y \oplus X \longrightarrow \mathcal{V}_{afa}[X]$, such that it holds:*

$$id \circ g = \wp_X \circ f$$

Proof: Obvious from the remarks in this section and Fact 14.1.1. q.e.d.

We finish our examination of fixed points with these remarks. Our next topic is the development of a generalized form of the corecursive substitution principle.

[4] Notice that the coding scheme was also very important in the development of non-well-founded set theory in [BarMo96]. The reader is referred to this book for further information.

14.3 Substitution

In this section, we give a generalized formulation of the corecursive substitution operation of Section 11.6. The crucial claim is formulated in Theorem 14.3.1. It shows that it is possible to extend a substitution function defined on a set of urelements X to a substitution operation defined on the whole universe $\mathcal{V}_{afa}[X]$.

A problem arises because the substitution function s of Section 11.6 is only specified with respect to a particular domain. The range of s remains underspecified. Here, we assume that the range of s is the class $Y \oplus \mathcal{V}_{afa}[Y]$ for a given collection of urelements $Y \subseteq \mathcal{U}$. Furthermore, we model substitution as an operation defined on a collection of urelements $X \subseteq \mathcal{U}$ and the universe $\mathcal{V}_{afa}[X]$. Hence, we do not assume that the universe is generated by the class of all urelements $\mathcal{U}$. This is a slight modification to the original version as developed in Section 11.6. In the following development, we assume that $\phi_X : \mathcal{V}_{afa}[X] \longrightarrow X \oplus \mathcal{V}_{afa}[X]$ is a final coalgebra with respect to $\wp_X$. Now, we state the theorem that claims that a substitution defined on urelements can be extended to a substitution operation defined on the universe.

Theorem 14.3.1 *Assume the anti-foundation axiom holds. Assume further that $s : X \longrightarrow Y \oplus \mathcal{V}_{afa}[Y]$ is a given substitution function. Then there exists a unique morphism $sub : \mathcal{V}_{afa}[X] \longrightarrow \mathcal{V}_{afa}[Y]$, such that the following diagram commutes.*

$$\begin{tikzcd}
\mathcal{V}_{afa}[X] \arrow[r, "\phi_X"] \arrow[d, "sub"'] & X \oplus \mathcal{V}_{afa}[X] \arrow[d, "\langle s, in_r \circ sub \rangle"] \\
\mathcal{V}_{afa}[Y] \arrow[r, "\phi_Y"'] & Y \oplus \mathcal{V}_{afa}[Y]
\end{tikzcd}$$

Proof: Intuitively, $s : X \longrightarrow Y \oplus \mathcal{V}_{afa}[Y]$ is the substitution function that substitutes urelements or sets for urelements taken from X. In a certain sense this substitution keeps the process running. Assume a function

$$r : (Y \oplus \mathcal{V}_{afa}[Y]) \oplus \mathcal{V}_{afa}[X] \longrightarrow Y \oplus (\mathcal{V}_{afa}[Y] \oplus \mathcal{V}_{afa}[X])$$

is given. r can be considered as a rearrangement.[5] We consider the following function e:

$$e \;=\; \langle r \circ (s \oplus id_{\mathcal{V}_{afa}[X]}) \rangle \circ \phi_X$$

Using a diagram e is specified as follows:

[5] This notion originates from [BarMo96]. Obviously r is an isomorphism.

$$\begin{tikzcd}
\mathcal{V}_{afa}[X] \arrow[r, "\phi_X"] \arrow[dr, "e"'] & X \oplus \mathcal{V}_{afa}[X] \arrow[d, "r \circ (s \oplus id_{\mathcal{V}_{afa}[X]})"] \\
& Y \oplus \mathcal{V}_{afa}[X \oplus Y]
\end{tikzcd}$$

Now, the premises of Fact 14.1.1 are satisfied and we can flatten the system. Hence, there exists a unique function $sub : \mathcal{V}_{afa}[X] \longrightarrow \mathcal{V}_{afa}[Y]$, such that the corresponding diagram in Fact 14.1.1 commutes. Formally, we get:

$$\phi_Y \circ sub \;=\; \langle id_{\mathcal{V}_{afa}[Y]}, sub \rangle \circ \langle r \circ (s \oplus id_{\mathcal{V}_{afa}[X]}) \rangle \circ \phi_X$$

An easy calculation shows that the following equality holds:

$$r \circ (s \oplus id_{\mathcal{V}_{afa}[X]}) \circ \phi_X \;=\; \langle s, in_r \circ sub \rangle \circ \phi_X$$

Hence, the diagram above commutes. Clearly, *sub* is unique, using the uniqueness of *sub* in Fact 14.2.4. q.e.d.

The above theorem is quite abstract. We will add some remarks in order to clarify the relation between the above theorem and the theory of non-well-founded sets.

Remark 14.3.1 (i) One can consider *sub* as an extension of s on sets. The unique morphism $\langle s, sub \rangle : X \oplus \mathcal{V}_{afa}[X] \longrightarrow Y \oplus \mathcal{V}_{afa}[Y]$ is the substitution operation as defined in Section 11.6 except that we skipped the identity on certain urelements.

(ii) The choice of the disjoint union in the above construction is appropriate for hypersets, because a coproduct construction is needed. Theorem 14.3.1 works also in a more general case where the requirement is only that $\oplus$ is an arbitrary coproduct operation.

(iii) Ordinary substitution works differently in comparison with corecursive substitution. Usually, one has base cases and a substitution operation defined on elements of sets. It is important to notice that ordinary substitution works step by step by substituting deeply embedded elements. This cannot be done similarly in the theory of hypersets, because hypersets do not have base cases in general and infinite descending chains of elements are not blocked as in the case of well-founded set theory.

(iv) The fact that the above theorem is formulated for a restricted universe $\mathcal{V}_{afa}[X]$ and not for the whole universe $\mathcal{V}_{afa}[\mathcal{U}]$ (where $X \subseteq \mathcal{U}$) is more a generalization than a restriction. Using the class $X = \mathcal{U}$ as urelements does not change anything in the theorem. Hence, Theorem 14.3.1 is general enough for applications.

We give two examples in order to clarify the ideas of the substitution principle as formulated in Theorem 14.3.1.

Example 14.3.2 (i) Assume a set $a = \{y, \Omega, \{a, x\}\}$ is given. Assume further that $s : \{x, y\} \longrightarrow \{\{\{1\}, \{1, 1\}\}, \{\{1\}, \{1, \{\omega, x\}\}\}\}$ is specified as follows:

$$s : x \longmapsto \langle 1, 1\rangle$$
$$s : y \longmapsto \langle 1, \{\omega, x\}\rangle$$

The substitution operation *sub* should result in the set

$$sub(a) = \{\langle 1, \{\omega, \langle 1, 1\rangle\}, \Omega, \{sub(a), \langle 1, 1\rangle\}\}$$

In order to show how Theorem 14.3.1 can be applied to the given set a, we specify the environment. First, notice that $X = \{x, y\}$ and $Y = \{x\}$. The substitution operation restricted to elements of X is specified by s. The operation $sub : \mathcal{V}_{afa}[X] \longrightarrow \mathcal{V}_{afa}[Y]$ is determined as above. *sub* is unique because of the construction examined in the proof of Theorem 14.3.1. In other words, *sub* is an extension of s to sets. Clearly, $\langle s, sub\rangle$ is the unique operation mapping $X \oplus \mathcal{V}_{afa}[X]$ into $Y \oplus \mathcal{V}_{afa}[Y]$ induced by the coproduct construction.

(ii) As a remark we mention that there is already the possibility to solve certain systems of equations. In order to see this, consider the following example. Assume a general system of equations $e : X \longrightarrow Y \oplus \mathcal{V}_{afa}[Y]$ is given, where $X = \{x_1, x_2, x_3, y_1, y_2\}$ and $Y = \{x_1, x_2, x_3, y_1, y_2\}$, such that the following conditions hold:

$$\begin{aligned} e(x_1) &= \langle 1, \{1, \{y_2\}, \Omega\}\rangle \\ e(x_2) &= \langle 1, \{\{y_2, x_1\}, y_1\}\rangle \\ e(x_3) &= \langle 1, \{0, \Omega\}\rangle \end{aligned}$$

An application of Theorem 14.3.1 guarantees that there is a unique substitution operation $\langle s, sub\rangle : X \oplus \mathcal{V}_{afa}[X] \longrightarrow Y \oplus \mathcal{V}_{afa}[Y]$, such that it holds:

$$\langle s, sub\rangle = \langle s, in_r \circ sub\rangle$$

This comes down to a unique substitution operation $\langle s, sub\rangle$ that extends the substitution defined on X to arbitrary sets $a \in \mathcal{V}_{afa}[X]$. With respect to elements $x \in X$ the substitution function s operates as e. In the following examples the extension of s to *sub* of the given function e works as follows:

$$\begin{aligned} sub(\langle 1, \{1, \{y_2\}, \Omega\}\rangle) &= \langle 1, \{1, \{y_2\}, \Omega\}\rangle \\ sub(\langle 1, \{\{y_2, x_1\}, y_1\}\rangle) &= \langle 1, \{\{y_2, \langle 1, \{1, \{y_2\}, \Omega\}\rangle\}, y_1\}\rangle \\ sub(\langle 1, \{0, \Omega\}\rangle) &= \langle 1, \{0, \Omega\}\rangle \end{aligned}$$

Notice that the substitution is not an ordinary substitution principle, because we work in a non-well-founded environment. There is (in general) no base case in order to start the substitution.

Remark 14.3.3 It seems to be the case that the consideration of the fixed point $X \oplus \mathcal{V}_{afa}[X]$ is a restriction, because there are no sets of the from $a = \{2, b\}$ in this universe. This is clearly misleading. First, the framework is a development in order to model systems of equations and therefore one needs a distinction between urelements (as non-sets) and sets. Second, there is an obvious isomorphism $f : \mathcal{V}_{afa}[X] \longrightarrow \{1\} \times \mathcal{V}_{afa}[X]$. Hence, nothing is lost using the disjoint union construction.

Notice that the difference between the corecursive substitution principle and the solution of a system of equations is simply the fact that in a system of equations it is not explicitly stated that e can map an element $x \in X$ into a set b, such that "x is somehow embedded in b". As a consequence one has to take a set b with $x \in b$ if one wants to apply the substitution principle to a system consisting of the equation $e : x \longmapsto \{x, a\}$. Clearly, b can be chosen as a set that obeys this condition, but that is not what we want precisely. We need a principle that gives us an account for precisely these systems where $X \subseteq \mathcal{U}$ and $e : X \longrightarrow Y \oplus \mathcal{V}_{afa}[Y]$.

In the next section, we shall consider the heart of the theory of non-well-founded set theory, namely the so-called corecursion theorem that was established originally in [Ac88]. This theorem can be seen first, as the dual of the recursion theorem and second, as a principle that guarantees that every system of set theoretical equations has a unique solution. We saw already different versions of this idea when we considered corecursive definitions and proofs by corecursion.

14.4 Corecursion

In this section, we will examine the corecursion theorem that goes back to [Ac88]. This theorem corresponds to the concept of a unique solution of a set theoretic system of equations on a very abstract level. The uniqueness of the solution is a consequence of the finality of an appropriate coalgebra. Fact 14.2.4 guarantees that arbitrary systems can be flattened, the substitution principle specified in Theorem 14.3.1 shows that we are able to establish a corecursive substitution principle even though non-well-founded sets are involved. The last step in the development of the theory of hypersets is the step that shows that there is a unique solution of systems of equations. The challenge is to find a unique solution of a general system of equations of the form $e : X \longrightarrow \mathcal{V}_{afa}[X \oplus Y]$. Notice that the indeterminates X do not occur in the solution-set of this system.

We assume in the following that for a given function $f : X \longrightarrow \mathcal{V}_{afa}[Y]$, the extension of f on sets is denoted by $[f] : \mathcal{V}_{afa}[X] \longrightarrow \mathcal{V}_{afa}[Y]$. The following theorem states the claim of the corecursion principle. As above we interpret the collections X and Y as collections of urelements.

Theorem 14.4.1 *Assume a general system of equations is given induced by the mapping $e : X \longrightarrow \mathcal{V}_{afa}[X \oplus Y]$. Provided that the anti-foundation axiom holds, there exists a unique solution morphism $sub : X \longrightarrow \mathcal{V}_{afa}[Y]$, such that it holds:*

$$sub = [(\langle in_r \circ sub, in_l \rangle)] \circ e$$

i.e. the following diagram commutes:

$$\begin{tikzcd}
X \arrow[r, "e"] \arrow[dr, "sub"'] & \mathcal{V}_{afa}[X \oplus Y] \arrow[d, "{[(\langle in_r \circ sub, in_l \rangle)]}"] \\
& \mathcal{V}_{afa}[Y]
\end{tikzcd}$$

Proof: Similarly to the proof of Theorem 14.3.1 we need an appropriate rearrangement. We introduce the following rearrangement r:

$$r : (X \oplus Y) \oplus \mathcal{V}_{afa}[X \oplus Y] \ \longrightarrow \ Y \oplus (X \oplus \mathcal{V}_{afa}[X \oplus Y])$$

Using r we can define the following morphism.

$$\wp_Y(r) \circ \phi_{X \oplus Y} : \mathcal{V}_{afa}[X \oplus Y] \ \longrightarrow \ \wp_Y(X \oplus \mathcal{V}_{afa}[X \oplus Y])$$

Using $\wp_Y(r) \circ \phi_{X \oplus Y}$ we can define a function

$$g : X \oplus \mathcal{V}_{afa}[X \oplus Y] \ \longrightarrow \ X \oplus \mathcal{V}_{afa}[X \oplus Y]$$

that generates a final coalgebra. Define g as follows:

$$g \ = \ \langle \wp_Y(r) \circ \phi_{X \oplus Y} \circ e, \wp_Y(r) \circ \phi_{X \oplus Y} \rangle$$

Obviously, $\langle X \oplus \mathcal{V}_{afa}[X \oplus Y], g \rangle$ is a final coalgebra. By finality there exists a function $s : X \oplus \mathcal{V}_{afa}[X \oplus Y] \longrightarrow \mathcal{V}_{afa}[Y]$ such that the following diagram commutes:

$$\begin{tikzcd}
X \oplus \mathcal{V}_{afa}[X \oplus Y] \arrow[r, "g"] \arrow[d, "s"'] & X \oplus \mathcal{V}_{afa}[X \oplus Y] \arrow[d, "\wp_Y(s)"] \\
\mathcal{V}_{afa}[Y] \arrow[r, "\phi_Y"'] & Y \oplus \mathcal{V}_{afa}[Y]
\end{tikzcd}$$

Define the substitution operation on systems of equations sub as follows: $sub = s \circ in_l$. In other words, the following diagram commutes:

$$\begin{tikzcd}
X \arrow[r, "in_l"] \arrow[dr, "sub"'] & X \oplus \mathcal{V}_{afa}[X \oplus Y] \arrow[d, "s"] \\
 & \mathcal{V}_{afa}[Y]
\end{tikzcd}$$

We claim that it holds: $[\langle in_r \circ sub, in_l \rangle] = s \circ in_r$.[6] In order to prove this claim notice that it holds:

$$(id \circ s) \circ r \;=\; \langle \langle in_r \circ s \circ in_l, in_l \rangle, in_r \circ s \circ in_r \rangle \;=\; \langle \langle in_r \circ sub, in_l \rangle, in_r \circ s \circ in_r \rangle$$

The single steps of the above equalities are easy to show using a diagram. Now, we can apply the functor $\wp_Y$ and we get the following equalities:

$$\begin{aligned}
\phi_Y \circ s \circ in_r &= \wp_X(s) \circ \wp_{X \oplus Y}(r) \circ \phi_{X \oplus Y} \\
&= \wp_X(\langle \langle in_r \circ sub, in_l \rangle, in_r \circ s \circ in_l \rangle) \circ \phi_{X \oplus Y}
\end{aligned}$$

Using Theorem 14.3.1 we can establish our claim, namely that it holds: $[\langle in_r \circ sub, in_l \rangle] = s \circ in_r$

It remains to show that it holds: $sub = [\langle in_r \circ sub, in_l \rangle] \circ e$. It is easy to see that it holds:

$$\phi_Y \circ [\langle in_r \circ sub, in_l \rangle] \circ e \;=\; \phi_Y \circ sub$$

Because mappings into final coalgebras are unique up to isomorphisms, we can deduce that it holds: $[\langle in_r \circ sub, in_l \rangle] \circ e = sub$. The uniqueness of sub is trivial.

[6] Notice that the injections are dependent on the context.

q.e.d.

Theorem 14.4.1 is quite abstract, but it fits nicely to the overall development. We mention some explanations in order to clarify the situation.

Remark 14.4.1 (i) The corecursion theorem provides an abstract notion of the principle that every general system of equations has a unique solution. The proof of the existence and uniqueness uses crucially the fact that $\langle \mathcal{V}_{afa}[Y], id_{\mathcal{V}_{afa}[Y]}\rangle$ is a final coalgebra. By finality there exists a unique morphism into this coalgebra. Notice that the correspondence between general systems and the coalgebraic modeling is quite close.

(ii) In our considerations, we used the (modified) power set functor $\wp_X$ as well as disjoint union of sets for the coproduct operation. It could be presumed that this functor is only a special case of the results in [BarMo96]. In fact, that is not the case. The reason for this is the fact that an arbitrary system of equations $e : X \longrightarrow \mathcal{V}_{afa}[X \oplus Y]$ is in fact as general as necessary. There are no inaccessible collections of $\mathcal{V}_{afa}[X \oplus Y]$, at least if one restricts the attention to all collections that can be built by the collection of urelements given by $X \oplus Y$.

(iii) Theorem 14.4.1 is called corecursion theorem, because a recursion-like property is described and specified without a foundation of this recursion. It turns out that the requirement that the anti-foundation axiom holds is essential in this context.

(iv) Notice that the anti-foundation axiom is needed in order to guarantee that $\langle \mathcal{V}_{afa}[X], id\rangle$ is a final coalgebras. Without finality of these objects the theorem cannot be proven. If we had no urelements but we would work in pure set theory, we could use the very same construction interpreting X and Y as parameters. That is the reason why Moss called his account parametric corecursion in [Mo∞b].

The following example shows how the corecursion theorem works practically. As a matter of fact, the theorem is simply an abstract version of the well-known fact that every general system of equations has a unique solution.

Example 14.4.2 Assume a system of equations is given by the following specification of the function e:

$$\begin{array}{rcl} e(x_1) & = & \{\{x_1\}, x_2, \emptyset\} \\ e(x_2) & = & \{\emptyset, 1, \omega\} \\ e(x_3) & = & \{y, \langle \Omega, x_3\rangle\} \end{array}$$

In our example, we use the set $X = \{x_1, x_2, x_3\}$ as collection of variables and $Y = \{y\}$ as a collection of further urelements with the property that $X \cap Y =$

$\emptyset$. Then, the above system can be represented as a morphism e, such that $e : X \longrightarrow \mathcal{V}_{afa}[X \oplus Y]$. The corecursion principle guarantees that there is a unique morphism $sub : X \longrightarrow \mathcal{V}_{afa}[Y]$, such that the following holds:

$$\begin{array}{rcl} sub(x_1) & = & \{\{sub(x_1)\}, sub(x_2)\} \\ sub(x_2) & = & \{0, 1, \omega\} \\ sub(x_3) & = & \{y, \langle \Omega, sub(x_3)\rangle\} \end{array}$$

The corecursion principle gives us precisely what we need to solve arbitrary systems of equations.

Some further remarks concerning the corecursion theorem and some additional ideas and facts in [Mo∞b] are given in the following consideration.

Remark 14.4.3 (i) We restricted our attention to the modified power set functor $\wp_X$. This is an appropriate functor concerning set theory, because in the constructible hierarchy every set in the universe can be reached using this functor. As a matter of fact, the developed construction is not dependent on this particular functor. Every functor which has final coalgebras can be used. For example, every functor Γ which is constructed using the operations $\{+, \times, (-)^a, \wp_{fin}, I\}$ has final coalgebras. The corecursion principle works similarly in this case.

(ii) In [Mo∞b], so-called *parametric corecursion systems* are introduced as the following six-tuple:

$$\mathfrak{C} = \langle \mathcal{C}, +, \Gamma, P, \pi, x \mapsto \langle x^*, \phi_x\rangle\rangle$$

Here, $\mathcal{C}$ is a category, $+$ is a coproduct operation, $\Gamma : \mathcal{C} \longrightarrow \mathcal{C}$ is an endofunctor, $\langle P, \pi\rangle$ is a final coalgebra, and finally the pair $\langle x, \psi_x\rangle$ with the property that $\psi_x : x \longrightarrow \Gamma(x + x^*)$ is a final coalgebra. In comparison with these parametric corecrecursion systems, we used the following systems:

$$\mathfrak{CLASS} = \langle \mathcal{CLASS}, +, \wp_X, \mathcal{V}_{afa}[X], id\rangle$$

We did not specify the pair $\langle x^*, \phi_x\rangle$ in our context but we used it implicitly. The obvious pairs resulting from the construction are the final coalgebras (non-well-founded universes relative to the generating urelements). It is important to notice that the whole development examined in this Chapter can be performed also in the framework of parametric corecursion systems.[7]

We add some remarks concerning the choice of appropriate functors that are used in [BarMo96].

[7] Cf. [Mo∞b].

Remark 14.4.4 (i) In [BarMo96], several kinds of functors are introduced. The most important type of functor in order to guarantee that the corecursion theorem works are so-called smooth operators. Intuitively, a smooth operator $\Gamma : \mathcal{C} \longrightarrow \mathcal{C}$ is a monotone operator with the properties that for all $a \subseteq Obj_{\mathcal{C}}$ the object $\Gamma(a) \subseteq \Gamma^*$ and the urelements of a do not interact with the urelements originating from Γ itself.

(ii) Although the introduction of different kinds of functors as developed in [BarMo96] has the advantage that a certain kind of fine structure can be examined, the present account is equally strong. The reason for this is the fact that we are able to consider uniquely arbitrary mappings into the final coalgebras.

(iii) The crucial fact behind the usage of uniform functors is the theorem that maximal fixed points of uniform functors are final coalgebras. This was examined by [Ac88], [BarMo96], and [Tu96].

We finish this chapter with these remarks. In the next chapter, we will apply the theory of non-well-founded sets to several problems stated in Chapter 2 of this work.

14.5 History

Most of the developments in this chapter are based on the paper [Mo∞b]. Fact 14.1.1 has a stronger version that was proven in [TuPl97]. Clearly, the principal idea that it should be possible to simplify the corecursion theorem as it was stated in [BarMo96] was obvious for a long time. There is a certain difference between the representations of [BarMo96], [Mo∞b] and this chapter. Whereas, in [BarMo96] every important fact is proven without using category theory explicitly, the representation in [Mo∞b] is a generalization of [BarMo96] without using urelements. The account developed here tries to give a precise representation of the theory in [BarMo96] in category theoretic terms including the usage of urelements. From a global perspective this chapter is somewhere between the representations in [BarMo96] and [Mo∞b] with specific differences to both alternative accounts.

Chapter 15

Applications of Hypersets

In the last chapter, we developed an abstract theory of non-well-founded sets. One motivation for the extension of the set theoretical universe was the difficulty to model circular phenomena in standard ZFC set theory. We saw in Part II and Part III of this work that it is possible to modify the underlying logic or the semantics (and definition theory) in order to get an account for circularity. Another possibility is to change the underlying set theoretic objects. In this chapter, we will examine the possibility to model circular phenomena by modifying set theory. The most prominent framework that uses hypersets is situation theory. In Section 15.1, we shall introduce the basic ideas of classical situation theory. We will give a coalgebraic reformulation of situation theory in Section 15.2, whereas applications of situation theory will be added in Section 15.3. The last section in this chapter is devoted to occurring problems, variations, and further remarks concerning situation theory.

15.1 Classical Situation Theory

15.1.1 Information Units

Classical situation theory can be introduced in different ways. An axiomatic approach can be found in [MoSel96]. We give a rather rough idea what is going one in situation theory without being completely explicit in our presentation. Later, we will introduce a more abstract account of situation theory using the theory of hypersets we developed in Chapter 14. We begin with some definitions introducing relational structures and the basic concept of infons.

Definition 15.1.1 *(i) A relational structure* $\mathcal{R} = \langle D, R_1^{n_1}, R_2^{n_2}, \ldots, R_m^{n_m} \rangle$ *is a* $n_m + 1$ *tuple where* D *is a given domain and each* $R_j^{n_j}$ *is a* n_j*-ary relation.*

(ii) A primitive infon σ *is a* $n_j + 2$*-tuple* $\sigma = \langle\langle R_j^{n_j}, a_1, a_2, \ldots, a_{n_j}, p \rangle\rangle$ *where* $R_j^{n_j}$ *is a relation of the relational structure* $\mathcal{R}$*. The arguments* a_i *are elements of the domain* D*.* $p \in \{0, 1\}$ *is called polarity of infon* σ *and can be identified with a negation in the case* $p = 1$*.*

Infons can be understood as elementary information units. The above definition is the most simplified version of an infon in situation theory. Notice that it is not possible for an information unit σ to contain another infon σ', nor is it possible that a more complex information unit can be embedded in an infon.[1] Last but not least, for a primitive infon it is not possible that there are 'holes' in it. For example, it is not possible that an argument place is not filled with an argument. In this respect, primitive infons are considered to be total. Later we will see that this notion can be generalized.

Remark 15.1.1 In order to develop a theory of infons, it is possible to require further restrictions. For example, one could add roles of arguments. This idea is linguistically motivated and has the advantage that argument positions are restricted for certain appropriate arguments. A further idea is to develop a hierarchical order of roles. Because our account is rather informative in this section, we do not complicate the situation by adding a spelled out theory of argument roles. But we assume that the linear order of the arguments is determined by the relation itself. Hence, the relation induces implicitly argument roles.

The following examples give an intuitive idea of the applications of infons in a linguistic environment.

Example 15.1.2 (i) Assume *cook* is a relation of arity 2. Furthermore, *Peter* and *dish* are given arguments. Consider the following infon σ:

$$\sigma \;=\; \langle\!\langle \mathit{cook};\ a_1\text{: } \mathit{Peter},\ a_2\text{: } \mathit{dish},\ 0\rangle\!\rangle$$

σ represents the information unit *Peter cooks a dish.* The linear order of the arguments is assumed to be determined by the relation *cook*. That means that the infon

$$\sigma' \;=\; \langle\!\langle \mathit{cook};\ a_1\text{: } \mathit{dish},\ a_2\text{: } \mathit{Peter},\ 0\rangle\!\rangle$$

represents the information unit *A dish cooks Peter.* Clearly, the interpretation in which the dish is the agens of the infon σ' is semantically nonsense (on a common sense level).

(ii) Consider the following infon τ:

$$\tau \;=\; \langle\!\langle \mathit{know};\ a_1\text{: } \mathit{Peter},\ a_2\text{: } \mathit{case},\ 1\rangle\!\rangle$$

The infon τ represents the sentence *Peter does not know the case.* Now, assume we do not want to use *case* for a particular proposition ϕ, say that *Oswald murdered Kennedy.* Then, we are faced with a problem, because the argument that needs to be filled in is a whole information unit. In our terminology, such

[1] We assume that D does not contain information units. Here, the elements of D are considered as objects without inner structure that can be used to construct information units, similarly to urelements in the case of set theory.

an information unit can be interpreted as an infon again. We want to express something like τ':

$$\tau' \;=\; \langle\!\langle \mathit{know};\ a_1\colon\ \mathit{Peter},\ a_2\colon\ \phi,\ 1\rangle\!\rangle$$

where $\phi \;=\; \langle\!\langle \mathit{murder};\ a_1\colon\ \mathit{Oswald},\ a_2\colon\ \mathit{Kennedy},\ 0\rangle\!\rangle$.[2] We conclude that we need an extension of the present approach in order to be able to substitute infons as arguments in a given infon.

We want to extend Definition 15.1.1(ii) of the concept of an infon to more complex information units. The following definition introduces a larger class of infons that is more appropriate for our context.

Definition 15.1.2 *Assume a relational structure $\mathcal{R}$ is given. An infon σ is a $n_j + 2$-tuple $\sigma \;=\; \langle\!\langle R_j^{n_j}, a_1, a_2, \ldots, a_{n_j}, p\rangle\!\rangle$ where $R_j^{n_j}$ is a relation of the relational structure $\mathcal{R}$. The arguments a_i are either the empty set, or elements of the domain D, or they are again infons σ'. $p \in \{0, 1\}$ is called polarity of infon σ and can be identified with a negation in the case $p = 1$. We call the collection of all infons INF^+.*

Some remarks concerning the generalization in the above definition are necessary.

Remark 15.1.3 (i) Notice that the above definition is not well-founded. We will see later that this does not create problems. Notice further that Definition 15.1.2 is not an inductive definition, because we did not specify a base case (or several base cases). Clearly, we could define all primitive infons as base cases. But as we will see in a reformulation of the above definition this does not suffice for our needs. In Definition 15.1.3, we will see the corecursive character of INF^+ more clearly.

(ii) The motivation for Definition 15.1.2 is the attempt to introduce infons that have the following form:

$$\sigma' \;=\; \langle\!\langle \mathit{know};\ a_1\colon\ \mathit{Peter},\ a_2\colon\ \langle\!\langle \mathit{murder};\ a_1\colon\ \mathit{Oswald},\ a_2\colon\ \mathit{Kennedy},\ 0\rangle\!\rangle,\ 1\rangle\!\rangle$$

Infon σ' is not circular. It is a perfectly well-founded information unit. But Definition 15.1.2 is general enough to define proper circular information units as well. For example, the following infon σ is not well-founded provided non-well-founded sets are available:

$$\sigma \;=\; \langle\!\langle \mathit{know};\ a_1\colon\ \mathit{Peter},\ a_2\colon\ \sigma,\ 0\rangle\!\rangle$$

This infon expresses the information σ, namely that Peter knows precisely this information σ. If the standard mathematical coding machinery of infons is set

[2] We do not consider tense features and possibilities to represent these features in situation theory. This extension is not in the focus of the present work.

theory, i.e. if one identifies infons with sets, then it is important to mention the fact that there is no set in ZFC that can code σ. Precisely at this point, non-well-founded set theory can be used. In ZFC_{afa}, there is a set that satisfies the properties of σ. The easiest way to guarantee the existence of σ is to consider the following set theoretic equation:[3]

$$x = \langle \texttt{know}, \texttt{Peter}, x, 0\rangle$$

Here, `know` and `Peter` can be considered as urelements. The above equation has a unique solution provided the anti-foundation axiom holds. This solution can be specified as follows:

$$s(x) = \langle \texttt{know}, \texttt{Peter}, s(x), 0\rangle$$

If we identify $s(x)$ with σ we generated a circular infon that satisfies the conditions determined by the equation $s(x) = \langle \texttt{know}, \texttt{Peter}, s(x), 0\rangle$.

(iii) A further remark is necessary in order to motivate the introduction of infons that contain 'holes'. From a naive perspective, it is reasonable to assume that an information unit of the form *Peter sees Mary playing the (...)* is a respectable information unit, although it is incomplete. For example, *Peter sees Mary playing the (...)* can be represented by the following infon.

$$\sigma \;=\; \langle\!\langle \textit{see};\ a_1\text{: } \textit{Peter},\ a_2\text{: } \langle\!\langle \textit{play};\ a_1\text{: } \textit{Mary},\ a_3\text{: } \emptyset,\ 0\rangle\!\rangle,\ 0\rangle\!\rangle$$

The effect of allowing holes in infons comes down to the possibility to represent partial information. Partiality becomes more and more important in linguistic theories. Examples of work concerning partiality in linguistics are [Mu89] in the context of event semantics in a four-valued logic and [PoSa94] in the context of HPSG. Partiality is also an important topic in DRT (cf. [KaRe93]).

It is obvious that our rough presentation of the general ideas of situation theory is not very precise. For example, we did not say anything concerning the governing principles of infons. Here, many choices are possible. For example, one could require that roles of arguments must satisfy certain consistency principles. We did not introduce roles explicitly, therefore it is not necessary to consider these questions. Another problem is how infons can be hierarchically ordered. An order of infons was the governing idea of the paper [BarEt90] where it was shown that certain restrictions on infons result in a complete Heyting algebra (or equivalently so-called frames in the sense of [Vic89]). Finally, an important point concerns the tools needed, in order to be able to identify certain infons. For example, if there are two infons σ and σ', then we would like to know whether the two infons are equivalent or not. Because we allow circular infons, ordinary equivalence relations will not be appropriate. In this context, bisimulations become important as the

[3] It is assumed that the set theoretic coding of the four-tuple $\langle \texttt{know}, \texttt{Peter}, x, 0\rangle$ is ordinary set theoretic coding of tuples.

relation that identifies structurally equal but not isomorphic objects. All these questions and challenges can be skipped for the moment. Certain questions will be considered later.

A last point should be mentioned. If one works with information units, it is sometimes quite important to merge two or more information units. For example, one would like to define an operation that takes as input two arbitrary infons and yields a complex infon that contains the two information units. This should be possible in a uniform way, even if the infons are different in their structure. More technically, the challenge is to develop a theory that defines a unification operation of heterogeneous information. It turns out that the development of such a theory is a hard problem. Certain accounts like the one in [ViVe96] uses a category theoretic construction, a so-called Grothendieck construction, for the merging operation of different information units. As far as the author knows, it is still not known whether it is possible to introduce a disjunction into this account. Other ideas were developed in [BarEt90] and in [BarSel97]. In the latter work, the authors develop a theory of constraints, namely so-called channel theory. The basic idea is that the change of information states is incorporated on a very elementary level in the theory. In channel theory, merging of information units can be interpreted as a coproduct operation of appropriate objects in a given category.

We will examine technical questions concerning the construction of complex infons later. Here, we only give a flavor of the problems of unification. In order to define an appropriate operation on infons, one has to assume certain properties of this connection. For example, one can assume that logical conjunction of infons is commutative: $\sigma \wedge \tau \ = \ \tau \wedge \sigma$. Unfortunately, this is in general not true for natural language. For example, sometimes a conjunction can express an order of the involved events as the following examples show.

(1) Peter walks to the church and listens to the words of the priest.
(2) Peter listens to the words of the priest and walks to the church.

Notice that (1) and (2) have different meanings. Hence, we can conclude that conjunction in natural language is not commutative.

Other possible principles that are candidates for governing conjunction are associativity and idempotence. Again, associativity is not trivially a correct principle for conjunctions in natural language. A discussion of this topic can be found in [Ha96]. In total, we see that it is quite difficult to figure out the formal properties of a merging operation of infons. Dependent on the context and the type of logical relations between simple infons one needs to choose different principles. We will come back to this topic later.

In the next subsection, we will consider certain ideas of how we can construct situations using infons. Situations can be interpreted as partial possible worlds. The idea that information states (situations) can be underspecified or in certain cases overspecified (hence, they can be classically inconsistent) is the leading idea of the construction.

15.1.2 Situations

In the above subsection, we presented the basic ideas of a theory of information units. The considerations were restricted to the possible information that can be expressed using infons. Furthermore, we considered which features infons should have when we try to model different types of information. We did not examine what it means if a certain infon holds (or is true) relative to a given information theoretic background. This subsection is devoted to this problem. First, we will begin with a definition clarifying the concept of a situation. Situations can be associated with a representation of partial information about the world.

We assume in the following that INF^+ is the collection of all infons relative to a given relational strcuture $\mathcal{R}$. We reformulate Definition 15.1.2 in a slightly modified form in order to emphasize the corecursive character of the definition.

Definition 15.1.3 *Assume a relational structure $\mathcal{R}$ is given. Then, $INF^+_{\mathcal{R}}$ is the largest collection, such that it holds:*[4] *if $\sigma \in INF^+_{\mathcal{R}}$ then $\sigma = \langle\langle R_i;\ a_1\ ,\ a_2\ ,\ \dots\ ,\ a_n\ ,\ p\rangle\rangle$, such that the following conditions (i) - (iii) hold:*

(i) $R_i \in \mathcal{R}$
(ii) $\forall a_j \in \{a_1, a_2, \dots, a_n\} : (a_j \in D \ \vee\ a_j = \emptyset \ \vee\ a_j \in INF^+_{\mathcal{R}})$
(iii) $p \in \{0, 1\}$

Notice that Definition 15.1.3 specifies a maximal fixed point. It is a coinductive definition and not an inductive definition. Notice that we did not introduce a merge operation as a unification of (possibly heterogeneous) infons in order to gain more complex infons. Last but not least, it is helpful to mention that we did not say anything about the ontological status of relations and elements of the domain D. Usually, it is assumed that these objects are urelements. We will consider this question in Section 15.2 more carefully.

It is clear that Definition 15.1.3 can only strengthen the expressive power in the case we work in non-well-founded set theory. In ZFC, there are no circular collections and therefore we would loose the intended interpretation. Because of this fact, we need to assume to work in ZFC_{afa} as our basic set theory. Although in alternative frameworks of circularity, as examined in Parts II and III of this work, we used classical set theory and changed the underlying logic or the semantical system, situation theory is based on classical logic. Circularity is introduced and modeled on the level of basic mathematical entities and not on a meta-level. That is the reason why it was necessary to develop the theory of hypersets.

[4] The subscript $\mathcal{R}$ is used to express that INF^+ is defined relative to a relational structure $\mathcal{R}$.

Our next definition clarifies the status of a situation. A situation can be understood as a collection of information units. It is clear that infons do not give us a complete picture of the world. What is needed is a procedure that tells us how situations interact with infons.

Definition 15.1.4 *(i) Assume a relational structure $\mathcal{R}$ is given. Then, a situation s is a collection of infons such that $s \subseteq INF^+$.*

(ii) The set of all situations is denoted by SIT.

Remark 15.1.4 (i) Notice that it is not assumed that a situation is consistent. It is allowed that a certain situation contains contradictory infons. For example, a situation $s \subseteq INF^+$ such that

$$\sigma \ = \ \langle\langle know;\ a_1\colon\ Peter,\ a_2\colon\ \phi,\ 0\rangle\rangle\ \in\ s$$

as well as

$$\sigma' \ = \ \langle\langle know;\ a_1\colon\ Peter,\ a_2\colon\ \phi,\ 1\rangle\rangle\ \in\ s$$

is not prohibited by Definition 15.1.4. This feature represents the possibility of an overspecified information state where inconsistent information units are available.

(ii) If we assume that the domain D of the given relational structure $\mathcal{R}$ is a set of urelements $X \subset \mathcal{U}$, then the collection of all infons INF^+ is again a set (clearly in $\mathcal{V}_{afa}[X]$).

In situation theory, the standard way to introduce the concept ϕ *is true in a situation* s is achieved by introducing a relation $\models\ \subseteq\ SIT \times INF^+_{\mathcal{R}}$. We did this implicitly, because $\models$ is defined via the elementhood relation as in Remark 15.1.4(i). In order to introduce this concept, we formulate the next definition.

Definition 15.1.5 *The support relation $\models\ \subseteq\ SIT\ \times\ INF^+_{\mathcal{R}}$ is defined as follows:*

$$\forall s \in SIT\, \forall \sigma \in INF^+ : (s\ \models\ \sigma\ \longleftrightarrow\ \sigma\ \in\ s)$$

The idea is that the expression $s \models \sigma$ represents that situation s supports infon σ. In other words, infon σ holds in situation s. Here, truth is introduced relative to a partial information state about the world. Hence, the idea is that only relative to a given information theoretic background an information unit can hold. We will see in Subsection 15.1.3 that it is possible to define propositions as pairs of a situation and an infon. Because of the fact that propositions are considered to be the bearers of truth, the importance of these pairs cannot be underestimated.

In the next subsection, we will add some remarks concerning complex information units and propositions.

15.1.3 Operations and Propositions

We would like to introduce some ideas concerning complex information units that are composed of simple information units. There are two problems we need to discuss roughly. First, we would like to give a semantical analysis of logically connected infons (for example, conjunctions of infons or disjunctions of infons). Second, we would like to give some ideas of a construction method of complex infons using heterogeneous simple infons. We begin with the first problem.

Definition 15.1.6 *(i) Assume a relational structure $\mathcal{R}$ is given. A situation s supports a conjunction of infons $\bigwedge \sigma_\alpha$ if and only if situation s supports all σ_β for $\beta \leq \alpha$. Formally we have:*

$$s \models \bigwedge \sigma_\alpha \quad \Leftrightarrow \quad s \models \sigma_\beta \quad \textit{for all } \beta \leq \alpha$$

(ii) Assume a relational structure $\mathcal{R}$ is given. A situation s supports a disjunction of infons $\bigvee \sigma_\alpha$ if and only if there exists an infon σ_β for $\beta \leq \alpha$, such that s supports σ_β. Formally we have:

$$s \models \bigvee \sigma_\alpha \quad \Leftrightarrow \quad s \models \sigma_\beta \quad \textit{for a } \beta \leq \alpha$$

Definition 15.1.6 is a straightforward application of standard logic to the theory presented so far. More difficult than disjunctions and conjunctions of infons is the treatment of negation of infons. Here, the problem arises that there are at least two possibilities to introduce a negation. One type of negation is already included in the basic definition of primitive infons. Because of the introduction of the polarity of an infon we have implicitly a negation on a very elementary level. The change of the polarity of an infon can be interpreted as a negation of that infon. There is a second possibility to express another type of negation, namely in the case when we take situations into account. Consider the following two expressions:

$$s \models \langle\langle \textit{be};\ a_1\!:\ \textit{Peter},\ a_2\!:\ \textit{in the kitchen},\ 1 \rangle\rangle$$

$$s \not\models \langle\langle \textit{be};\ a_1\!:\ \textit{Peter},\ a_2\!:\ \textit{in the kitchen},\ 0 \rangle\rangle$$

Can we require that these two propositions are equivalent? Clearly, in total consistent situations they are, but in situations with underspecified information this equivalence is not appropriate. If situation s does not contain any information about Peter, then s does not support that Peter is in the kitchen, but this does not imply that s supports that Peter is not in the kitchen. Here, matters become more complicated. A similar situation arises in overspecified situations. It can be possible that s supports that Peter is in the kitchen *and* that Peter is not in the kitchen if there is too much information available. A discussion concerning different forms of negation in situations theory can be found in [BarEt87].

The only property we can specify concerning negated infons is a special case of Definition 15.1.5: Assume a relational structure $\mathcal{R}$ is given. Then, the following holds:

$$s \models \langle\langle R_i;\ a_1\ ,\ a_2\ ,\ \ldots\ ,\ a_n\ ,\ 1\rangle\rangle \Leftrightarrow \langle\langle R_i;\ a_1\ ,\ a_2\ ,\ \ldots\ ,\ a_n\ ,\ 1\rangle\rangle \in s$$

We conclude that negation requires a more sophisticated discussion in situation theory, because of the partiality of the account and different possibilities to introduce a negation.

We need to consider the second problem mentioned above. In [MoSel96], a first account of a unification of infons is proposed. Unfortunately, the development is not very explicit. Especially, it is not clear how it is possible to introduce disjunctions of infons. Simplifying matters, the idea in [MoSel96] is roughly the following one. First, we introduce an order relation $\sqsubseteq$ on infons, then we introduce a relation $ConOf \subseteq INF^+ \times INF^+$, and finally we define the conjunction $\sigma_1 \wedge \sigma_2$ of simple infons σ_1 and σ_2 as the least infon τ in the $\sqsubseteq$ order, such that $ConOf(\sigma_1, \tau)$ and $ConOf(\sigma_2, \tau)$ holds.[5] In other words, the conjunction of σ_1 and σ_2 is the least infon τ, such that σ_1 is a conjunct of τ and σ_2 is a conjunct of τ.

The following definition makes this idea more precise by introducing a coding machinery that helps keeping the involved distinct objects distinct. On the one hand we need to speak about atomic (non-structured) objects and on the other hand we need to code structured objects (infons).

Definition 15.1.7 *Let INF^+ be a collection of infons relative to a extensional structure $\mathcal{R}$. We define $B_0 = \{0\} \times (|\mathcal{R}| - INF^+)$ and $B_1 = \{1\} \times \wp(INF^+)$. Finally define for each $a \in |\mathcal{R}|$:*

$$a^* = \begin{cases} \langle 0, a\rangle & : \ if\ a\ \notin\ INF^+ \\ \langle 1, \{a\}\rangle & : \ if\ a\ \in\ INF^+ \end{cases}$$

Then, we can define a new structure $\mathcal{B}$ with $|\mathcal{B}| = B_0 \cup B_1$ according to the following condition.

$$\forall R : R(a_1^*, a_2^*, \ldots, a_n^*) \in \mathcal{B} \quad \Leftrightarrow \quad R(a_1, a_2, \ldots, a_n) \in \mathcal{R}$$

Concerning the properties of $ConOf$ we define:

$$ConOf\langle\langle 1, \tau\rangle, \langle 1, \sigma\rangle\rangle \quad \Leftrightarrow \quad \tau \subseteq \sigma$$

Remark 15.1.5 (i) Notice that the defined conjunction is commutative, associative, and idempotent. Intuitively, B_0 codes atomic objects, i.e. objects that are not infons, whereas B_1 codes the collection of all structured objects, namely infons.

[5] The term *simple infon* is used to express that simple infons are used to build more complex ones. It has nothing to do with the technical meaning of simple infons in [MoSel96].

(ii) In [MoSel96], a further possibility is mentioned to introduce a conjunction. This particular conjunction is not commutative, but associative and idempotent. There are other possible variations.

(iii) A problem arises, if one wants to model disjunctions in a similar way as conjunctions. It is not sufficient simply to define a disjunction $\sigma_1 \vee \sigma_2$ as the greatest infon τ in the $\sqsubseteq$ order, such that $DisOf(\sigma_1, \tau)$ and $DisOf(\sigma_2, \tau)$ holds where $DisOf$ is defined as follows:

$$DisOf\langle\langle 1, \tau\rangle, \langle 1, \sigma\rangle\rangle \quad \Leftrightarrow \quad \sigma \subseteq \tau$$

Why is this problematic? As a consequence of Definition 15.1.7 we get the following equality for conjunctions:

$$\langle 1, \sigma\rangle \wedge \langle 1, \tau\rangle \ = \ \langle 1, \sigma \cup \tau\rangle$$

Although this is appropriate for conjunction, the dual equality concerning disjunctions makes no real sense. An expression of the form $\langle 1, \sigma\rangle \vee \langle 1, \tau\rangle \ = \ \langle 1, \sigma \cap \tau\rangle$ does not result necessarily in an infon of the form $\langle 1, x\rangle$. Here the ideas of forming complex infons can only partially be applied and further work needs to be done to develop a theory of structured objects.

A last ingredient is necessary to get a more complete picture of situation theory. We need to introduce propositions in order to get a version of truth-functional semantics. The following definition specifies the concept of a proposition.

Definition 15.1.8 *Assume a relational structure $\mathcal{R}$ and a collection INF^+ of infons and a collection SIT of situations are given. Then, an expression p is a proposition if and only if p is of the form $s \models \sigma$ or $s \not\models \sigma$ where $s \in SIT$ and $\sigma \in INF^+$.*

Remark 15.1.6 (i) Definition 15.1.8 is important for the semantics of situation theory, because the bearers of truth are propositions in the so-called Austinian version of situation theory. This differs quite significantly from the development in Parts II and III of this work where the bearers of truth were sentences. Clearly, We will discuss this feature of situation theory in Chapter 16 more closely.

(ii) In [BarEt87], the definition of a proposition differs slightly from Definition 15.1.8. Whereas in Definition 15.1.8 we introduced a proposition as a certain relation between a situation and an infon, in [BarEt87] a proposition is a relation between a situation and a type of an infon. Clearly, the latter account is more precise, but for our purpose Definition 15.1.8 suffices.

(iii) The standard set theoretic coding of a given proposition $p \ = \ (s \models \sigma)$ can be achieved by interpreting p as a pair:

$$p \;=\; (s \models \sigma) \;=\; \langle s, \sigma \rangle$$

Clearly, there are alternatives to represent a proposition set theoretically. For example, one could imagine to represent a proposition as an infon of the following form

$$p \;=\; \langle\!\langle \models;\ a_1\colon\ s,\ a_2\colon\ \sigma,\ 0 \rangle\!\rangle$$

Using $\models$ as a structural relation that is not included in the relations given by the relational structure $\mathcal{R}$ one can interpret p as an infon again. But we will use the classical representation of p as a pair.

The fact that only propositions can be true or false in the Austinian picture of situation theory makes it necessary to extend our development in order to be able to represent circular propositions. We want to introduce objects of the form $p \;=\; (s, \langle\!\langle \mathbf{T};\ a_1\colon\ p,\ 0 \rangle\!\rangle)$ that are interpreted as the proposition p that claims that p is true in s. The following definition makes this idea precise.

Definition 15.1.9 *Assume a relational structure $\mathcal{R}$ and the collection $INF^+_{\mathcal{R}}$ (relative to $\mathcal{R}$) is given. $PROP$, $INF^+_{\mathbf{T}}$, and SIT^+ are defined as the largest classes such that it holds:*

- *If $\sigma \in INF^+_{\mathbf{T}}$, then σ satisfies condition (i), condition (ii), or condition (iii):*

  (i) $\sigma \in INF^+_{\mathcal{R}}$
  (ii) $\sigma = \langle\!\langle \mathbf{T};\ a_1\colon\ p,\ 0 \rangle\!\rangle$
  (iii) $\sigma = \langle\!\langle \mathbf{T};\ a_1\colon\ p,\ 1 \rangle\!\rangle$

- *If $p \in PROP$, then p satisfies one of the conditions (iv)-(vi):*

  (iv) $p \;=\; (s, \sigma)$ for $s \in \wp(INF^+)$ and $\sigma \in INF^+$
  (v) $p \;=\; (s, \langle\!\langle \mathbf{T};\ a_1\colon\ p,\ 0 \rangle\!\rangle)$ for $s \in SIT^+$, $p \in PROP$, and $\mathbf{T} \in \mathcal{U}$
  (vi) $p \;=\; (s, \langle\!\langle \mathbf{T};\ a_1\colon\ p,\ 1 \rangle\!\rangle)$ for $s \in SIT^+$, $p \in PROP$, and $\mathbf{T} \in \mathcal{U}$

- *Every $s \in SIT^+$ is a subset of $INF^+_{\mathbf{T}}$*

Remark 15.1.7 (i) We interpret the truth predicate $\mathbf{T}$ as an urelement in conditions (ii), (iii), (v), and (vi) above. Clearly, that is in accordance with our development so far.

(ii) The fact that we are working in the non-well-founded universe guarantees that circular propositions of the form $p \;=\; (s, \langle\!\langle \mathbf{T};\ a_1\colon\ p,\ 0 \rangle\!\rangle)$ for $s \in SIT^+$ and $p \in PROP$ do exist. p expresses that p itself is true in s. It is important to notice that there are different Truth-teller propositions dependent on the particular situation s. This models the intuition that Truth-teller propositions are interpreted relative to a particular information theoretic background. The same is true for Liar-like propositions. Relative to a particular situation one can construct (using the anti-foundation axiom) a proposition

$$f_s \ = \ (s, \langle\!\langle \mathbf{T};\ a_1\colon\ f_s,\ 1\rangle\!\rangle)$$

The dependency on situations is the reason why Truth-teller and Liar propositions are usually denoted by f_s and t_s in [BarEt87] emphasizing the particular situation a proposition is referring to. Moreover, this makes clear that in situation theory the context is a constitutive and not an additional feature.

(iii) Notice that Definition 15.1.9 is a corecursive definition. This fits nicely into our coalgebraic picture of non-well-founded set theory.

We finish this section with these remarks. In the next section, we will use our more abstract framework of parametric corecursion in order to give a representation of situation theory.

15.2 A Coalgebraic Version of Situation Theory

In this section, we give a coalgebraic reformulation of the central ideas of situation theory. First, we need to introduce primitive infons, then we will give an account of circular infons. Finally, some remarks will be added concerning the representation of situations and models. We begin our considerations with the definition of primitive infons.

Definition 15.2.1 *Assume a relational structure $\mathcal{R} = \langle D, R_1^{n_1}, R_2^{n_2}, \ldots, R_m^{n_m}\rangle$ is given. We consider the domain D as well as the relations $R_i^{n_i}$ as a collection of urelements, such that $D \cap \{R_1^{n_1}, R_2^{n_2}, \ldots, R_m^{n_m}\} = \emptyset$. We define primitive infons similarly as in Definition 15.1.1(ii).*

Remark 15.2.1 We assume that the collection of all primitive infons INF is a set. Elements of the domain D are considered as objects without any inner structure. This is not a necessary condition, but it simplifies matters. We assume further (as usual) that a proper class of urelements $\mathcal{U}$ is given in order to be able to speak about the real world.

We need to define the set INF^+ (as defined in Definition 15.1.2) using primitive infons. We want to use the corecursion principle in order to construct this collection. The following fact makes it precise in which way non-well-founded set theory and corecursion are involved in this development.

Fact 15.2.2 *Assume the collection of all primitive infons INF is given. Assume further that the modified power set functor $\wp_X$ is given where $X \subseteq \mathcal{U}$ is a set of urelements. Consider an arbitrarily specified morphism*

$$e : X \longrightarrow \mathcal{V}_{afa}[X \oplus (D \cup \{R_1^{n_1}, R_2^{n_2}, \ldots, R_m^{n_m}\})]$$

Then, there exists a unique morphism

$$sub : X \longrightarrow \mathcal{V}_{afa}[(D \cup \{R_1^{n_1}, R_2^{n_2}, \ldots, R_m^{n_m}\}]$$

given by the corecursion theorem. Furthermore, INF^+ is a subset of the collection of all solution sets of the above construction.

Proof: The proof is an easy application of the corecursion theorem. q.e.d.

Fact 15.2.2 is a simple application of the corecursion theorem 14.4.1. We add some remarks concerning the features of the collection of infons.

Remark 15.2.2 (i) Notice that the usage of the disjoint union makes it superfluous to require that X is new for $D \cup \{R_1^{n_1}, R_2^{n_2}, \ldots, R_m^{n_m}\}$. This simplifies matters significantly. On an abstract level, the reason for the complicated definition of appropriate functors (for example proper functors, smooth functors etc.) in [BarMo96] is a direct result of using products instead of coproducts.

(ii) The morphism $e : X \longrightarrow \mathcal{V}_{afa}[X \oplus (D \cup \{R_1^{n_1}, R_2^{n_2}, \ldots, R_m^{n_m}\})]$ maps urelements into the (non-well-founded) set theoretical universe based on the collection $D \cup \{R_1^{n_1}, R_2^{n_2}, \ldots, R_m^{n_m}\}$ of urelements. Considering plainly set theory this is completely sufficient to guarantee the existence of circular infons. The problem is that a solution of an arbitrary function e cannot necessarily associated with an infon. For situation theory we need a more specific range of e, namely a range where the elements are (finite) sequences of the form $\langle\langle R_i;\ a_1\ ,\ a_2\ ,\ \ldots\ ,\ a_n\ ,\ p\rangle\rangle$. In Definition 15.2.4, we give an alternative account for the generation of infons where we get closer to this requirement.

(iii) From a theoretical point of view, the existence and uniqueness of the morphism *sub* guarantees that every appropriate infon can be modeled in the present account.[6] Notice that we need to assume that the anti-foundation axiom holds. Without the anti-foundation axiom the maximal fixed point $\mathcal{V}_{afa}[X \oplus (D \cup \{R_1^{n_1}, R_2^{n_2}, \ldots, R_m^{n_m}\})]$ would not contain circular sets. Hence, the corecursion theorem does not *generate* non-well-founded sets, but guarantees that *sub* is unique.

The following fact states that every infon σ can be reached by the above construction.

[6] Clearly, *sub* generates for arbitrary functions e elements that are not of the correct form.

Fact 15.2.3 *For every infon $\sigma \in INF^+$ there is a morphism*

$$e : X \longrightarrow \mathcal{V}_{afa}[X \oplus (D \cup \{R_1^{n_1}, R_2^{n_2}, \ldots, R_m^{n_m}\})$$

such that the solution morphism $sub : X \longrightarrow \mathcal{V}_{afa}[D \cup \{R_1^{n_1}, R_2^{n_2}, \ldots, R_m^{n_m}\}]$ of Fact 15.2.2 contains σ.

Proof: The claim is an obvious consequence of the construction of circular infons in Section 15.1 and the properties of the solution mapping *sub*. q.e.d.

In Remark 15.2.2(ii) we mentioned the necessity to adjust Fact 15.2.2, in order to get a better approximation of the possible solutions to the intended infons and their structure. The next definition makes this idea precise.

Definition 15.2.4 *Assume a relational structure $\mathcal{R} = \langle D, R_1^{n_1}, R_2^{n_2}, \ldots, R_m^{n_m}\rangle$ is given. We define a functor Γ as follows:*

$$\Gamma : S \longrightarrow \bigcup_{0 \leq n \leq \omega} (\{R_1^{n_1}, R_2^{n_2}, \ldots, R_m^{n_m}\} \times (S \oplus D)^n \times \{0,1\})$$

Definition 15.2.4 mirrors the structure of infons. Notice that circular infons are a proper subset of the collection of all infons generated by Definition 15.2.4. The following proposition ensures that Γ has a maximal fixed point and that the resulting solutions of an arbitrary function e satisfying certain conditions (namely mapping a collection of urelements into a the maximal fixed point of Γ), define a subset of the intended collection of all infons INF^+.

Proposition 15.2.5 *(i) Assume Γ is the functor specified in Definition 15.2.4. Then, Γ induces a final coalgebra $\langle \Gamma^*, \alpha_{\Gamma^*}\rangle$ where Γ^* is taken to be the maximal fixed point of Γ.*

(ii) Assume that the anti-foundation axiom holds. If e is an arbitrary morphism with the property

$$e : X \longrightarrow (\bigcup_{0 \leq n \leq \omega} (\{R_1^{n_1}, R_2^{n_2}, \ldots, R_m^{n_m}\} \times (X \oplus Y \oplus D)^n \times \{0,1\}))^*$$

then there exists a unique morphism sub, such that

$$sub : X \longrightarrow (\bigcup_{0 \leq n \leq \omega} (\{R_1^{n_1}, R_2^{n_2}, \ldots, R_m^{n_m}\} \times (Y \oplus D)^n \times \{0,1\}))^*$$

and sub satisfies: $sub = (\langle in_r \circ sub, in_l\rangle) \circ e$.

Proof: (i) The construction of Γ uses the Cartesian product, disjoint union, and exponents. All these operations preserve the existence of final coalgebras according to Fact 13.4.3. Notice that in the above case it is not longer necessary to work in $\mathcal{CLASS}$. Even in $\mathcal{SET}$ the existence of the final coalgebra for Γ is guaranteed.

(ii) According to claim (i), the pair $\langle \Gamma^*, \alpha_{\Gamma^*} \rangle$ is the final coalgebra for Γ with Γ^* to be the maximal fixed point. Applying the corecursion theorem 14.4.1, the claim is immediately justified. q.e.d.

Remark 15.2.3 (i) Proposition 15.2.5 shows that it is possible to give an account in an extremely general setting. The solutions of the system of equations induced by the morphism e provides circular infons (together with other infons). This is the abstract version of situation theory we developed in Section 15.1.

(ii) Clearly, even in Proposition 15.2.5 the morphism e could possibly map an urelement $x \in X$ into a set $e(x)$, such that $sub(x)$ does not correspond to an intended infon. This is a direct consequence of the generality of the approach, because the resulting collections are set theoretic universes build on certain urelements.

In Section 15.1, we introduced not only circular infons but also circular propositions. Propositions were modeled by pairs of the form $p = \langle s, \sigma \rangle$ where s is a given situation and σ is a given infon. It is clear that the construction in Fact 15.2.2 suffices to guarantee that circular propositions exist. This is stated in the following fact.

Fact 15.2.6 *Assume an urelement* $\mathbf{T} \in D$ *is given. For every proposition* p *there is a morphism*

$$e : X \longrightarrow \mathcal{V}_{afa}[X \oplus (D \cup \{R_1^{n_1}, R_2^{n_2}, \ldots, R_m^{n_m}\})]$$

such that p *is contained in the range of the solution morphism* $sub : X \longrightarrow \mathcal{V}_{afa}[D \cup \{R_1^{n_1}, R_2^{n_2}, \ldots, R_m^{n_m}\}]$.

Proof: Obvious. q.e.d.

Notice that the construction in Proposition 15.2.5(ii) does not suffice to represent propositions as solutions of system of equations. But as should be clear by the considerations so far it is easy to extend this construction to fit into the overall picture. We will not go into details concerning this point.

What can be said concerning complex infons? Complex infons are a problem in situation theory. A possibility to introduce complex infons is to use category theoretic constructions. The intuitive idea is to associate a conjunction of infons with a coproduct operation. This is the usual way how these connectors are modeled. In Subsection 15.1.3 we saw that for complex infons the following association is possible.

$$\langle 1, \sigma \rangle \wedge \langle 1, \tau \rangle \ = \ \langle 1, \sigma \cup \tau \rangle$$

In order to store the information where σ and τ came from, it is better to use a coproduct operation $\oplus$ instead of ordinary union. It is clear that this construction can easily be incorporated in the present framework. As is obvious from the considerations in Definition 15.1.7 and Remark 15.1.5, there are various ways to model conjunctions with different properties (commutativity, associativity etc.). Using coproducts for conjunctions limits the flexibility of the particular modeling. In this respect, the general approach is not as flexible as the more elementary approach.

In Section 15.1, we defined situations as (possibly inconsistent) collections of infons. In the coalgebraic version of situation theory, this definition can be adopted. The important impact of the theory of coalgebras is the possibility to define circular infons (and circular propositions) on a very abstract and general level. Hence, we consider the definitions to be applicable to the coalgebraic framework as well.

In the next section, we will discuss several applications and examples of situation theory. The examples mentioned in Chapter 2 will be in the focus of our consideration.

15.3 Applications of Situation Theory

Standard applications of situation theory are the modeling of truth and pathological expressions on the one hand, and the modeling of knowledge on the other. A further application are perception reports. We will examine the first and the second application, because they deal explicitly with circularity and revision processes. Representations of perception reports will not play any role in this section.

15.3.1 Truth and Situation Theory

The following list specifies the main differences between theories of truth we examined in Parts II and III of this work and the ideas constituting a theory of truth in the context of situation theory.

- The bearers of truth in situation theory are propositions and not sentences like in the Gupta-Belnap systems or in Kripke's account.

- Situation theory tries to give an account to model the behavior of pathological propositions. The motivation is not to define a truth predicate. In this respect, situation theory is similar to revision theories, but differs significantly from Kripke's account.

- Situation theory incorporates the context in which a sentence is uttered (in difference to the account of Gupta-Belnap and Kripke). The idea that every proposition is a pair of a situation and an information unit is crucial here. In this respect, situation theory provides a more relativistic theory for pathological expressions.

- In general, situation theory is an account from a more information theoretic perspective in comparison with the alternative accounts.

Remark 15.3.1 It is clear that these differences are not independent of each other. For example, a strong argument for choosing propositions as bearers of truth (instead of sentences) is the intuitively correct claim that the interpretation of sentences is highly dependent on the context. Hence, context sensitivity is important if one wishes to argue for the claim that propositions are the bearers of truth. On the other hand, if we choose propositions to be the bearers of truth it is clear that the theory needs features to represent a context. Obviously, we have dependencies between these claims.

We will examine certain properties of situation theory in several examples. First, we will consider the classical examples of pathological propositions like the Liar proposition and the Truth-teller proposition.

Example 15.3.2 (i) We begin our examination of examples with the Truth-teller. First, we need a formal representation of this expression. The standard idea is to use an urelement **T** as truth predicate. Naively, the information the Truth-teller sentence codes (without any context) can be expressed according to the following information unit:

$$\sigma \;=\; \langle\!\langle \mathbf{T};\ \sigma,\ 0 \rangle\!\rangle$$

σ expresses the information that this very information σ is in the extension of **T**. Although σ is a circular infon it is impossible for σ to have a truth value, because only propositions can be true or false (in the Austinian picture). Hence, what is needed is a context. Consider the following proposition:

$$\emptyset \;\models\; \langle\!\langle \mathbf{T};\ \sigma,\ 0 \rangle\!\rangle$$

According to the definition σ does not hold in $\emptyset$, because it holds $\sigma \notin \emptyset$. What can be said about a proposition in which s includes σ? Consider the following example:

$$s \;=\; \{\ldots,\sigma,\ldots\} \;\models\; \langle\!\langle \mathbf{T};\ \sigma,\ 0 \rangle\!\rangle$$

In this context it holds $\sigma \in s$. Naively, this is a kind of Truth-teller like expression. But we do not express a Truth-teller proposition, because a Truth-teller claims something about a proposition, not about an infon (information unit). Additionally, there is no connection between σ and s except the fact that $\sigma \in s$. We need to refer to circular propositions in order to represent the Truth-teller proposition. Consider the following proposition:

$$p_s \;=\; (s, \langle\langle \mathbf{T};\ a_1\colon\ p_s,\ 0\rangle\rangle) \qquad \text{for } s \in SIT^+$$

This proposition claims that this proposition is true in situation s. What can be said about its truth conditions? Dependent on s the Truth-teller p_s can be true or false. For example, if $s = \emptyset$, p_s is clearly false. To see that there are true Truth-teller propositions, assume s is an arbitrarily given situation. Construct s' as follows: $s' = s \cup \{\langle\langle \mathbf{T};\ a_1\colon\ p_{s'},\ 0\rangle\rangle\}$. Then, $p_{s'}$ is obviously true. Hence, there are situations in which the Truth-teller is true and there are situations in which the Truth-teller is false. This is in accordance to our intuitions.

(ii) We want to give an analysis of the Liar proposition. The Liar can be represented as follows:

$$\lambda_s \;=\; (s, \langle\langle \mathbf{T};\ \lambda_s,\ 1\rangle\rangle)$$

λ_s is the proposition that claims its own falsity in situation s. As a matter of fact it is clear that λ_s can be false. For example, if s is the empty set or a situation in which only well-founded information units are collected the Liar proposition in s is false. But λ_s can also be true. For example, if s is an arbitrary situation consider $s' = s \cup \{\langle\langle \mathbf{T};\ \lambda_{s'},\ 1\rangle\rangle\}$. Then, the Liar proposition $\lambda_{s'}$ is true. Notice that we do not absolutely claim anything about the truth-value of Liar proposition. We make a claim about the Liar proposition relative to a particular situation (information theoretical background).

The Liar proposition is intuitively interpreted as a proposition that has an instable behavior if one tries to evaluate it. After a particular truth value is assumed for the Liar proposition every further step in the reasoning about the Liar results in a change of its truth value. We want to represent this behavior with situation theoretic tools. The following example discusses this point.

Example 15.3.3 Assume the Liar proposition

$$\lambda_s \;=\; (s, \langle\langle \mathbf{T};\ a_1\colon\ \lambda_s,\ 1\rangle\rangle)$$

is given. We want to model the reasoning about the Liar. First, λ_s makes a statement about s, namely that it is false in s. Assume that λ_s is false. We extend situation s to the situation $s_1 = s \cup \{\langle\langle \mathbf{T};\ a_1\colon\ \lambda_s,\ 1\rangle\rangle\}$ like in Example 15.3.2(ii). According to our semantics we have developed so far the following proposition is in fact true:

$$p_s = (s_1, \langle\langle \mathbf{T};\ a_1\colon\ \lambda_s,\ 1\rangle\rangle)$$

This represents the evaluation after the first revision in the reasoning about the Liar proposition. The next step can be represented as follows.

$$\lambda_{s_1} = (s_1, \langle\langle \mathbf{T};\ a_1\colon\ \lambda_{s_1},\ 1\rangle\rangle)$$

Clearly this proposition is false. But a further revision yields a situation $s_2 = s_1 \cup \{\langle\langle \mathbf{T};\ a_1\colon\ \lambda_{s_1},\ 1\rangle\rangle\}$. Then the proposition

$$p_{s_1} = (s_2, \langle\langle \mathbf{T};\ a_1\colon\ \lambda_{s_1},\ 1\rangle\rangle)$$

is true. It is clear how the reasoning can be continued to arbitrarily many circles. This representation models quite nicely how we reason about the Liar proposition.

The examples show that pathological propositions can be modeled in situation theory. Because of the fact that propositions are the bearers of truth, we can make several distinctions dependent on the particular perspective. But it seems to be the case that in our modeling the truth and falsity of pathological propositions do not depend on the facts in world. It seems that we don't need models of the world at all in order to analyze pathological expressions. We saw in Chapter 2 that it is quite often the case that pathological propositions are dependent on facts in the world. In the remaining part of this subsection we will add some comments concerning models of the world in situation theory.

First, we need to introduce models representing the facts that hold in the world. The intuitive idea of models is to interpret them as consistent situations. The following definition introduces partial and total models.

Definition 15.3.1 *(i) A (partial) model $\mathfrak{M}$ is a collection of infons, such that the following conditions hold:*

- *If $\sigma \in \mathfrak{M}$ then $\neg\sigma \notin \mathfrak{M}$.*
- *$\langle\langle \mathbf{T};\ a_1\colon\ p,\ 0\rangle\rangle \in \mathfrak{M}$ iff p is true in $\mathfrak{M}$*
- *$\langle\langle \mathbf{T};\ a_1\colon\ p,\ 1\rangle\rangle \in \mathfrak{M}$ iff p is false in $\mathfrak{M}$*

(ii) A total model $\mathfrak{M}$ is a partial model, such that $\mathfrak{M}$ is not properly contained in any model $\mathfrak{M}'$.

Remark 15.3.4 (i) We introduced the concept of truth of a proposition as a relation between situations and infons. Truth of a proposition relative to a model $\mathfrak{M}$ introduces a further aspect: The proposition must be contained in a consistent set of infons. Hence, not every arbitrary collection of infons can support a proposition p if we wish to guarantee the truth of p in model.

(ii) The introduction of two concepts of truth, namely one that simply refers to an information theoretic background and a second one that refers additionally to a model seems to be quite counterintuitive. As a matter of fact, this is only a problem at first sight. The distinction is plausible if one wants to model real world phenomena. Here, often the situation occurs that a proposition is true in a particular context, whereas this proposition does not hold in an appropriate model of the world. In a certain sense, the history of sciences seems to represent precisely this situation.

We summarize some facts about models in situation theory. These facts will give some more intuitions about the situation theoretic account.

Proposition 15.3.2 *(i) In a total model $\mathfrak{M}$ the Truth-teller proposition*

$$p_s \;=\; (s, \langle\!\langle \mathbf{T};\ p_s,\ 0\rangle\!\rangle)$$

can be true in $\mathfrak{M}$ and can be false in $\mathfrak{M}$.

(ii) In a total model $\mathfrak{M}$, a Liar proposition cannot be true in $\mathfrak{M}$.

(iii) In a total model $\mathfrak{M}$ the Truth-teller circle can be true in $\mathfrak{M}$ and can be false in $\mathfrak{M}$.

Proof: (i) Clearly, the Truth-teller proposition p_s can be false in $\mathfrak{M}$. In order to see that the Truth-teller can be also true, assume s is any situation such that it holds: $\forall\sigma \in s : \sigma \in \mathfrak{M}$. Then, extend s to s' similar to the construction in Example 15.3.2(i), namely by extending s with the appropriate Truth-teller proposition $p_{s'}$. Because $p_{s'}$ is true it is an element of $\mathfrak{M}$.

(ii) For a Liar proposition λ_s that is true, it holds $\langle\!\langle \mathbf{T};\ a_1{:}\ \lambda_s,\ 1\rangle\!\rangle \in s$. If $s \subseteq \mathfrak{M}$, then λ_s is false in $\mathfrak{M}$ according to the definition of truth in $\mathfrak{M}$.

(iii) This can be proven similarly to claim (i). q.e.d.

We add some remarks to the above considerations, in order to relate these properties to the alternative accounts in Parts II and III of this work. Although we will present a more extensive examination concerning the relations between the different accounts in Chapter 16, it is useful to mention some of the main differences already here.

Remark 15.3.5 (i) Proposition 15.3.2(ii) shows that in a total model $\mathfrak{M}$ of the world the Liar proposition cannot be true in $\mathfrak{M}$. But the claim is even stronger. Because of the fact that situation theory does not use a multi-valued logic these propositions are plainly false. Here, an important difference between accounts like the ones presented in Parts II and III and situation theory arises. Kripke interprets pathological sentences as entities that are different from true sentences and false sentences. In Gupta-Belnap systems, these sentences are false, but by reconsidering the behavior of them in the revision process they can be distributed into different classes. In situation theory, we have simply false propositions from a global perspective. Finer distinctions can be made by considering the behavior of these propositions in particular situations. We saw in Example 15.3.3 that it is possible to represent the unstable behavior of the Liar proposition as well. The framework of situation theory seems to be quite flexible concerning this point.

(ii) Notice that the Truth-teller circle was a problem in the Gupta-Balnap account (at least in the standard theory). In situation theory, the treatment of this circle is straightforward.

(iii) Although situation theory can provide a theory of truth it is important to realize that situation theory does not generate a truth predicate in the classical sense. The truth concept in situation theory is (because of its context sensitivity) more relativistic in comparison with the alternative theories of truth. Only situation theory is a framework that uses an information theoretic background as an important relativization. The considered alternative accounts are more classical. Further remarks concerning the relation between situation theory on the one hand, and Gupta-Belnap systems and Kripke's account on the other can be found in Chapter 16.

In the next subsection, we will consider knowledge representation systems and the possibility to model knowledge in situation theory. We will examine some of the puzzles used in Chapter 2 in order to show how situation theory can be successfully applied in this context.

15.3.2 Knowledge Representation

A theory of knowledge representation tries to code information units interpreted as elements of a contextually given information background in a formal framework. For our purpose, it is not necessary to develop a theory of data structures. We will simply represent data in situation theory. In particular, we are interested how the modeling of the difference between common ground and private knowledge can be achieved.

We refer the reader to Chapter 2 concerning examples of the difference between common ground and private knowledge. The representation of private knowledge is relatively straightforward. If a person a knows a fact ϕ,[7] then we can represent this fact by the expression $\Box_a\phi$. Such a representation is simple and completely according to our intuitions. The question is: What does it mean that a proposition ϕ is common ground? There are several proposals for an analysis of this concept. We begin with the proposal by Schiffer in [Sc72]:

Definition 15.3.3 *A proposition ϕ is common ground for a group of discourse members $a_1, a_2, \ldots, a_n$ if and only if the following two conditions hold:*[8]

(i) $\forall i \in \{0, 1, \ldots, n\} : \Box_{a_i}\phi$
(ii) $\forall i \in n \forall j \in n : \Box_{a_i}\phi \longrightarrow \Box_{a_j}(\Box_{a_i}\phi)$

[7] At this point, we do not distinguish facts, propositions, and sentences from each other.

[8] In this definition, we do not identify propositions with the technical meaning of propositions in situation theory like the one we used in Subsection 15.3.1.

Remark 15.3.6 (i) Notice that the above definition has an inductive character. Because of the fact that expressions of the form $\Box_{a_i}\phi$ are again considered as propositions, an induction process yields sequences of simple propositions that are known by a particular discourse member. These sequences can have arbitrary finite length.

(ii) Our definition requires that i and j are natural numbers. Nothing prevents us (except the intuition) to extend Definition 15.3.3 to arbitrary ordinals. We restrict our considerations to sequences of finite length. The smallest set that satisfies the conditions of the definition above is the collection of all finite chains of iterated knowledge operators.

Schiffer's definition is not the only possibility to represent common ground. Another analysis was proposed by David Lewis in [Le69]. Using situation theoretic terms the following definition represents David Lewis' definition.[9]

Definition 15.3.4 *A proposition ϕ is common ground for a group of discourse members $a_1, a_2, \dots a_n$ if and only if the following $n+1$ conditions hold:*

$$\begin{array}{l}
s \models \phi \\
s \models \langle\langle know;\ a_1,\ s,\ 0\rangle\rangle \\
s \models \langle\langle know;\ a_2,\ s,\ 0\rangle\rangle \\
\vdots\ \ \vdots\ \ \vdots\ \ \vdots\ \ \vdots\ \ \vdots \\
s \models \langle\langle know;\ a_n,\ s,\ 0\rangle\rangle
\end{array}$$

Remark 15.3.7 (i) Notice that Lewis' definition is circular. In order to specify the situation s, one has to know the infons that constitute s. In order to determine the infons, one is forced to know s again. Except the first requirement all conditions are circular propositions. Without the usage of hypersets there would not exist a set that can represent one of these infons (except the first one). In this respect, hypersets are necessary to represent Lewis' ideas.

(ii) Clearly, another formulation of the circular conditions above would be the following one:

$$\forall i \in \{1, 2, \dots, n\} : s \models \langle\langle know;\ a_i,\ s,\ 0\rangle\rangle$$

We used the explicit formulation in order to make the idea more transparent.

We mention a last analysis of common ground that goes back to [Ha77]. We specify this analysis in the next definition.

[9] The following formulation goes back to [Bar90].

Definition 15.3.5 *A proposition ϕ is common ground for a group of discourse members $a_1, a_2, \ldots, a_n$ if and only if the following condition holds:*

> *The information $\tau = [\forall i \in \{1, 2, \ldots, n\} : \Box_{a_i}(\phi \wedge \tau)]$ is known by each member a_j for $j \in \{1, 2, \ldots, n\}$.*

Remark 15.3.8 (i) Similarly to Definition 15.3.4 the above definition is circular: The information unit τ contains τ itself as an element. From the pure definition it is not clear whether Definition 15.3.5 and Definition 15.3.4 are equivalent or not.

(ii) From a conceptual point of view it seems to be the case that Definition 15.3.5 is more difficult to understand than Definitions 15.3.4 and 15.3.3. The reason for this is probably the fact that every definitional property is packed into one formula.

First, we show that all definitions of common ground can be reformulated in situation theoretic terms. This is important to guarantee that situation theory is a tool strong enough to model what we want. Fact 15.3.6 shows that all presented definitions of common ground can be reformulated in situation theoretic terms.

Fact 15.3.6 *(i) Definition 15.3.3 can be represented in situation theory.*
(ii) Definition 15.3.5 can be represented in situation theory.

Proof: (i) Consider Definition 15.3.3. We translate both conditions into situation theory terms. This can be done straightforwardly: the following two conditions specify precisely these two conditions.

$$\begin{aligned} \forall i \in n : \ s &\models \langle\langle know;\ a_i,\ \phi,\ 0\rangle\rangle \\ \forall k \in n \forall j \in n : \ s &\models \langle\langle know;\ a_j,\ \phi,\ 0\rangle\rangle \\ \Longrightarrow s &\models \langle\langle know,\ a_j,\ \langle\langle know;\ a_k,\ \phi,\ 0\rangle\rangle,\ 0\rangle\rangle \end{aligned}$$

The above formulas are literal translations of the conditions in Definition 15.3.3. That suffices to show the claim.

(ii) Consider Definition 15.3.5. As in (i) we give a literal translation of the conditions in that definition.

$$\begin{aligned} &\forall i \in n : \ s \models \langle\langle know;\ a_i,\ s',\ 0\rangle\rangle \\ &\forall i \in n : \ s' \models \langle\langle know;\ a_i,\ \phi,\ 0\rangle\rangle \wedge \langle\langle know;\ a_i,\ s',\ 0\rangle\rangle \end{aligned}$$

Again the above conditions are obviously equivalent to the conditions in Definition 15.3.5. This shows the claim. q.e.d.

Remark 15.3.9 Notice that in the above fact we split one condition into two conditions using two situations s and s'. This simplifies the understanding of the complex relation. As a matter of fact it is also possible to express the same in one formula. Then, the readability is quite complicated.

In [Bar90], some interesting results concerning a comparison of the different possibilities of modeling common ground are proven. We mention two important facts here.

Fact 15.3.7 *(i) If we restrict our attention to finite situations s, then a countable approximation of s in the sense of Definition 15.3.3 is equivalent to the circularly defined situation according to Definition 15.3.5.*

(ii) If we assume that all situations are finite, then Definition 15.3.4 and Definition 15.3.5 are equivalent.

Proof: Compare [Bar90] for further information. q.e.d.

In order to make the quite abstract considerations concrete, we consider the following example. We will model the muddy children problem using situation theoretic techniques. The problem with the muddy children problem is to represent the reasoning of the children. In other words: We want to model the logic behind the reasoning that is used by the children in order to figure out whether they are muddy or not.

Example 15.3.10 Consider the muddy children problem (16) in Section 2.3. In order to keep things readable, we restrict the number of children to 2 (more precisely we consider the children a_1 and a_2) and we assume that both children are dirty on the forehead. We consider the reasoning of a_1. After the first question of the father, a_1 knows situation s_1 where s_1 supports the following infons.[10]

$$(1) \quad s_1 \models \langle\langle know;\, a_1,\, \langle\langle dirty;\, a_1,\, 0\rangle\rangle \vee \langle\langle dirty;\, a_2,\, 0\rangle\rangle,\, 0\rangle\rangle$$

$$(2) \quad s_1 \models \langle\langle know;\, a_1,\, \langle\langle dirty;\, a_2,\, 0\rangle\rangle,\, 0\rangle\rangle$$

(1) is justified because of the statement of the father that at least one child is dirty. (2) is justified because a_1 sees that a_2 is dirty. Additionally to propositions (1) and (2) known by a_1, the following two constraints hold and are known by both children a_1 and a_2.

$$(3) \quad s_1 \models (\langle\langle know;\, a_2,\, \langle\langle dirty;\, a_2,\, 0\rangle\rangle,\, 0\rangle\rangle \Longrightarrow \langle\langle say;\, a_2,\, \langle\langle dirty;\, a_2,\, 0\rangle\rangle,\, 0\rangle\rangle)$$

[10] We simply the notation slightly: We assume that the infons supported by situations s_1 and s_2 are known by a_1.

$$(4)\quad s_1 \models (\langle\langle know;\, a_2,\, \langle\langle dirty;\, a_i,\, 0\rangle\rangle,\, 0\rangle\rangle \\ \wedge\ \langle\langle know;\, a_2,\, \langle\langle dirty;\, a_1,\, 1\rangle\rangle,\, 0\rangle\rangle) \\ \Longrightarrow \langle\langle know;\, a_2,\, \langle\langle dirty;\, a_2,\, 0\rangle\rangle,\, 0\rangle\rangle$$

These specifications of situation s_1 (after the first question of the father) are sufficient to represent the reasoning of a_1 after the second question of the father. After the second question a_1 reasons as follows. From (3) a_1 can infer (5) using contraposition:

$$(5)\quad s_2 \models \langle\langle say;\, a_2,\, \langle\langle dirty;\, a_2,\, 0\rangle\rangle,\, 1\rangle\rangle \\ \Longrightarrow \langle\langle know;\, a_2,\, \langle\langle dirty;\, a_2,\, 0\rangle\rangle,\, 1\rangle\rangle$$

From (5) a_1 infers (6) using (4):

$$(6)\quad s_2 \models \langle\langle know;\, a_2,\, \langle\langle dirty;\, a_2,\, 0\rangle\rangle,\, 1\rangle\rangle \\ \Longrightarrow \langle\langle know;\, a_2,\, \langle\langle dirty;\, a_1,\, 1\rangle\rangle,\, 1\rangle\rangle$$

Because a_1 knows that it holds: $s_2 \models \langle\langle say;\, a_2,\, \langle\langle dirty;\, a_2,\, 0\rangle\rangle,\, 1\rangle\rangle$, using modus ponens two times a_1 can deduce (7) from (5) and (6):

$$(7)\quad \langle\langle know;\, a_2,\, \langle\langle dirty;\, a_1,\, 1\rangle\rangle,\, 1\rangle\rangle$$

Moreover, a_1 knows (8), because a_1 assumes that a_2 does not have a visual defect:

$$(8)\quad \langle\langle know;\, a_2,\, \langle\langle dirty;\, a_1,\, 0\rangle\rangle,\, 0\rangle\rangle \vee \langle\langle know;\, a_2,\, \langle\langle dirty;\, a_1,\, 1\rangle\rangle,\, 0\rangle\rangle$$

Obviously, the following two infons are contradictory:

$$\langle\langle know;\, a_2,\, \langle\langle dirty;\, a_1,\, 1\rangle\rangle,\, 1\rangle\rangle \\ \langle\langle know;\, a_2,\, \langle\langle dirty;\, a_1,\, 1\rangle\rangle,\, 0\rangle\rangle$$

Hence, a_1 can deduce (knows) that (9) holds:

$$(9)\quad s_2 \models \langle\langle know;\, a_2,\, \langle\langle dirty;\, a_1,\, 0\rangle\rangle,\, 0\rangle\rangle$$

Finally, from (9) it follows (10), assuming that knowledge of a fact implies that this fact holds (which is generally accepted). Hence, a_1 can deduce (10):

$$(10)\quad s_2 \models \langle\langle dirty;\, a_1,\, 0\rangle\rangle$$

That represents the reasoning of a_1. Clearly, the reasoning of a_2 is precisely the dual reasoning. The application of both instances of reasoning yields the result that after two questions of the father, both children say that they are dirty at the forehead.

Remark 15.3.11 (i) Our presentation of the logical reasoning of the children is slightly simplified because of the number of involved children. In general, this reasoning can be extended to an arbitrary finite number of children. Clearly, the representation of situations become quite complicated by a larger number of involved children.

(ii) We introduced several constraints in the above reasoning, like the constraint that if a_2 knows that a_2 is dirty, then a_2 is immediately saying that a_2 is dirty. A further constraint is that a_1 knows that a_2 knows whether a_1 is dirty or not. We assume here that neither a_1 nor a_2 has a visual defect.

(iii) We mentioned several other examples of the difference between common ground and private knowledge. It is clear that the examples are structurally similar to the muddy children problem. The differences in form can be easily adjusted.

There is an important insight from the modeling in Example 15.3.10. A representation of common ground that uses neither inductive principles like in Schiffer's analysis (compare Definition 15.3.3) nor circular principles like in Harman's analysis (compare Definition 15.3.5) and in Lewis definition (compare Definition 15.3.4) cannot be successfully applied to the muddy children problem. This is true because any fixed finite number of iterations of the knowledge operators is insufficient for modeling the phenomenon if there are $m > n$ children involved and $n + 1$ many children are dirty on their forehead. Hence, in order to guarantee a successful modeling either an inductive approach or a circular approach is necessary.

In the next section, we mention some major problems of situation theory that are known for a long time.

15.4 Problems and Further Remarks

Historically, situation theory was developed in order to model constraints in natural language that cannot be represented in classical Montague semantics. The successful modeling of certain phenomena ranging from circularity of knowledge representation to the development of a theory of propositions has also certain disadvantages. The following list summarizes these disadvantages.

(1) There is no known model for situation theory.

(2) How can modal logic be introduced?

(3) Situation theory gives no reliable account for a theory of argument roles.

(4) The original aim to develop a theory of constraints for natural and formal languages was not achieved.

We need to add some general remarks concerning the above points. After these remarks we will check whether these problems are challenging the possibility to apply situation theory to circular phenomena.

Remark 15.4.1 (i) The problem to find a model for situation theory is a relatively old problem. Years of work were spent in order to fix this problem. There is still no solution. Clearly, from a conceptual point of view this situation is quite unsatisfactory, but applications can nevertheless be successfully represented (as one can see in this chapter). If one works with strong and clearly specified restrictions, then it is possible to find a model for particular applications. This is obvious from Definition 15.1.9 and the applications of this definition in the above parts of this chapter.

(ii) In situation theory, there is still no idea how to introduce modal operators. In the further development of situation theory, namely in channel theory Barwise developed an account in which modal operators can be introduced (compare [Bar97] for further information). Unfortunately, channel theory is not a direct extension of situation theory, although the two authors Barwise and Seligman of the monograph [BarSel97] point out explicitly that channel theory should be considered as such an extension. To examine the relation between situation theory and channel theory precisely remains a task for future research and is not spelled out yet.

(iii) An old problem in linguistics is the development of a theory of argument roles. Situation theory has the same problems as alternative semantical theories. There is no generally accepted solution for this problem. As is shown in [MoSel96] there are first steps towards formalizing (and axiomatizing) minimal conditions that are required for a theory of argument roles. An interesting question is whether it is possible to incorporate classical work about argument structure in the linguistic realm into the formal framework of situation theory.[11]

(iv) Interestingly enough, the original motivation for the development of situation theory, namely the modeling of the problem how constraints in natural language can be modeled in a formal framework, was not solved by the theory. Although more than ten years were spent to develop the theory, this problem is still open. The reason for this is perhaps that the classical features of situation theory prevent a successful development of a theory of constraints. Channel theory was able to solve this problem (at least partially). The price one has to pay is that channel theory is no longer a classical logical framework but rather a category theoretic framework that is in general not complete and not sound.

[11] For example, [Gr90] can count as a theory that provides a linguistic framework for a theory of argument structures. The question is whether such theories can be incorporated into situation theory.

We want to check whether these problems have a significant influence on the modeling of circularity using situation theoretic tools. As we saw in Section 15.3, situation theory can be used to represent pathological expressions in natural language and the distinction between private knowledge and common ground quite successfully. Furthermore, situation theoretic techniques can be used to compare different analyses of the common ground puzzles. Only the representation of circular information units using hypersets guarantees the possibility to prove that the presented inductive account and the circular accounts are equivalent. In this respect, it is clear that the above problems situation theory faces are not an a priori argument against the usage of situation theory for the modeling of circular phenomena.

Furthermore, it is clear that the only real challenge of situation theory is problem (1). The application of situation theory to circularity does not presuppose modal operators, a theory of argument roles, nor an explicit and spelled out theory of constraints. The fact that there is no model for situation theory becomes a problem if one wants to develop a general theory of circularity, applicable to all kinds of circularity ranging from a theory of truth for natural language (not only for a fragment of natural language), to circular phenomena in knowledge representation, mathematics, computer science etc. Clearly, then situation theory is not sufficient: Without the possibility to decide whether there is a consistent set of infons or not we are facing a non-trivial problem.

It is relatively straightforward to assume that situation theory can be applied to various other phenomena. This is true because of the flexibility of the theory itself. For example, it should be possible to give a situation theoretic modeling of circularity in computer science. The representation of labeled transition systems as introduced in Definition 13.1.1 can be represented as a model where the infons constituting that model are elements σ of the following form

$$\sigma \;=\; \langle\langle \longrightarrow_S;\ s,\ a,\ s',\ 0\rangle\rangle$$

Clearly, that needs to be spelled out, but we do not want to go into details concerning this point. In general, we can say that there are further possibilities of applications of situation theory with respect to the modeling of non-well-founded phenomena.

We finish this chapter with these remarks. The final chapter of this work summarizes the considerations and gives an outlook of future work and alternative frameworks that were not in the focus of this work.

15.5 History

Situation theory was originally developed by Jon Barwise and John Perry in [BarPe83] and [Bar81]. A further development of situation theory can be found in [De91]. Another influence was the introduction of non-well-founded sets into situation theory in [BarEt87] and (theoretically) in [Ac88]. In [Ki94], this line of development was further investigated. A readable version of situation theory can be found in [MoSel96]. The applications are folklore to the subject and were the main motivation for the development of situation theory. The attempt to find a model for situation theory is documented in several papers like [Ac90, Ac96] and [Lu91, AcLu91]. The problem of defining complex infons goes back to [BarEt90]. Applications of situation theory as presented here can be found in [BarEt87] concerning pathological expressions and [Bar90] concerning knowledge representation. As far as the author knows the explicit development of the logic of the muddy children problem is new here. Alternative accounts of modeling the muddy children problem (common ground versus private knowledge) can be found in [Gr94] (using dynamic logic), and [BaMoSo∞] (using an extended form of modal logic).

Chapter 16

Final Remarks

It is time to look back at the concepts we have examined so far. In this work, we considered a variety of different frameworks that were originally developed in order to model circularity. Most of them were developed directly as a theoretical answer to the challenges of paradoxes and the pathological behavior of certain philosophical and linguistic concepts. In this chapter, we will try to give ideas for further developments that could lead to new results in this subject. We will also add some further comments concerning the relations between and a comparison of the proposed frameworks. Although they are different in nature, they have many similarities (clearly on different levels).

We will begin with remarks concerning a comparison of the different frameworks. In this respect, complexity theoretic aspects come into consideration as well as the (sometimes implicit, sometimes explicit) mathematical and philosophical assumptions. Clearly, we do not absolutely rank the frameworks with respect to some suspect categories concerning their quality. Each theory has its advantages and its disadvantages. Furthermore, it seems to be obvious that the continuous work in modeling circularity yields more advanced and, from a general perspective, better frameworks.

16.1 A Comparison of the Different Frameworks

In this section, we want to compare the three different frameworks. In some sections above, we already have given some ideas as to how to compare Kripke's account with the account developed by Gupta and Belnap. Now, we will try to extend these considerations to the third theory developed in Part IV as well. It is clear that for every framework there was a motivation. Kripke developed his account of partial defined truth predicates in order to provide a theory of truth. A quite similar motivation can be specified for the revision theoretic account by Herzberger, Gupta, and Belnap. And the situation theoretic account using non-well-founded sets originates from questions concerning knowledge representation and paradoxes. Although it turns out that some of the further developments were motivated by other possible applications, it is clear that one cannot reasonably require that Kripke's account should solve the problem of how to model the private knowledge vs. common ground distinction. That makes it quite difficult to compare the frameworks.

First of all, we examine the complexity of the systems. We spend a lot of effort to determine the complexity of Gupta-Belnap systems in Chapter 8. The Kripke approach as well as the approach using non-well-founded sets is much more straightforward.

16.1.1 The Complexity of the Frameworks

As we examined in Subsection 9.3.3, it is possible to define Kripke's fixed point approach in the context of revision theories. This is possible because Kripke's inductive approach can be modeled in the expressively stronger Gupta-Belnap systems (at least in the semantical systems $\mathbf{S}^{\#}$ and $\mathbf{S}^{*}$). We give some examples.

Example 16.1.1 (i) Assume a language L is given. A Kripke-style fixed point approach with respect to a three-valued logic guarantees the existence of fixed points. Consider the minimal fixed point d. Clearly, one can reach this point by an inductive definition, starting with the bottom element $\perp$ of possible evaluations and iterating the applications of a monotone operator $\Gamma : \{T, F, N\}^{Sent_{L^+}} \longrightarrow \{T, F, N\}^{Sent_{L^+}}$. This process results in a fixed point. It is obvious that this is an inductive definition. Informally, the following definition specifies the minimal fixed point using an inductive definition (where $\mathfrak{M}$ denotes the classical ground model and $\mathfrak{M}'$ denotes the standard model of arithmetic). The minimal fixed point $\perp : Sent_{L^+} \longrightarrow \{T, F, N\}$ is the smallest set S, such that the following condition holds:

$$\begin{aligned}
\forall \phi \in Sent_L \forall \psi \in Sent_{L^+} : \quad & ([[\phi]]^{\mathfrak{M}} = T) \ \rightarrow \ \langle \phi, T \rangle \in S \\
\wedge \ & ([[\phi]]^{\mathfrak{M}} = F) \ \rightarrow \ \langle \phi, F \rangle \in S \\
\wedge \ & ([[\phi]]^{\mathfrak{M}} = N) \ \rightarrow \ \langle \phi, N \rangle \in S \\
\wedge \ & ([[\phi]]^{\mathfrak{M}'} = T) \ \rightarrow \ \langle \phi, T \rangle \in S \\
\wedge \ & ([[\phi]]^{\mathfrak{M}'} = F) \ \rightarrow \ \langle \phi, F \rangle \in S \\
\wedge \ & ([[\psi]]^{\mathfrak{M}}) = T) \ \rightarrow \langle \mathbf{T}(\psi), T \rangle \in S
\end{aligned}$$

Because of the fact that the semantical systems $\mathbf{S}^{\#}$ and $\mathbf{S}^{*}$ can model arbitrary inductive definitions,[1] the minimal fixed point in a Kripke-style approach can be represented in revision theories.

(ii) In Kripke's construction, there are in general several maximal fixed points. These points can also be defined using the semantical systems $\mathbf{S}^{\#}$ and $\mathbf{S}^{*}$. One has to take into consideration that similar to the case of inductive definitions, there is the possibility to define coinductive definitions revision theoretically as well.[2] In this respect, Kripke's account has less expressive strength in comparison with revision theories.

[1] Compare Lemma 8.4.2.

[2] Compare Corollary 8.4.3.

(iii) We need to consider the maximal intrinsic fixed point for a moment. According to Kripke, this is an important point, in order to construct a truth predicate in the object language (at least in one possible reading of [Kr75]). This particular fixed point can be defined as the infimum of all maximal fixed points.[3] Because the maximal fixed points are revision theoretical definable and the infimum relation is arithmetical, the maximal intrinsic fixed point can be defined in $\mathbf{S}^*$ and $\mathbf{S}^{\#}$.

Remark 16.1.2 (i) Dependent on the choice of particular fixed points the complexity of truth in a Kripke-style approach is either Π^1_1 or Σ^1_1. The classical construction using the minimal fixed point is a Π^1_1 construction. This is a direct consequence of Example 16.1.1.

(ii) We developed the modeling of Kripke's account using revision theoretic techniques in a three-valued logic. It is clear that it is also possible to develop his account in a four-valued logic. The situation in a four-valued (monotone) logic is even simpler, because there is no maximal intrinsic fixed point (in the general case).

The complexity theory of revision theories was developed in Sections 8.3 and 8.4. We summarize the facts from these sections.

- The complexity of validity of $\mathbf{S}^{\#}$ is Π^1_2.
- The complexity of validity of $\mathbf{S}^*$ is Π^1_2.
- The complexity of the definable subsets of ω is at least $\Pi^1_2 \cup \Sigma^1_2$.

As a direct consequence of these considerations, we can state that revision theories are stronger concerning their expressive power in comparison to classical Kripke-style accounts. The truth concepts of these two approaches are therefore quite different. This is further supported by the fact that the theories show quite different behaviors in certain applications.[4]

The last framework we examined in this work was the theory of non-well-founded sets. In particular, situation theory was used (extended with non-well-founded sets) in order to give an account of a theory of truth and a theory of knowledge representation. The examination of the complexity of non-well-founded set theory is more complicated, because we work in a different setting from the very beginning: we changed the underlying set theory. Furthermore, the axiomatization we formulated in Section 11.1 contains a non-standard technique, namely the claim that there is a proper class of urelements. We did not

[3] Compare Theorem 4.4.2(iii).

[4] Compare Section 10.1 for further explanations concerning this point.

introduce an axiomatization of classes, but refrained from such an axiomatization non-well-founded set theory can be axiomatized in a first-order style like classical ZFC set theory.

If one considers the techniques we used developing a theory of hypersets as well as the introduction of situation theory, then it is clear that the usage of corecursive methods plays an important and crucial role. Notice that the usage of the final coalgeba in order to identify the universe of non-well-founded sets, as well as the collection of infons that can be used in situation theory, makes it impossible to characterize the theory by Π^1_1 formulas. What we need are Σ^1_1 definitions.

We summarize the above considerations in the following list specifying the properties of the non-well-founded set theory account:

- Non-well-founded set theory is first-order definable, provided the existence of a proper class of urelements is given.
- The definition of the universe of non-well-founded set theory is a Σ^1_1 statement.
- The definition of the collection of infons that are used in our context has complexity Σ^1_1.

In the next subsection, we shall consider the mathematical assumptions of the different frameworks.

16.1.2 Mathematical Assumptions

There is an obvious difference between the considered frameworks. Whereas in the case of Kripke's account and the revision theoretic approach, classical ZFC set theory is used (without urelements) this is no longer true for the situation theoretic account where essentially non-well-founded sets are used. Notice that the introduction of non-classical set theory changes the framework on an elementary level. In non-well-founded set theory it is possible to speak about objects that do not exist in ordinary mathematics which is usually based on ZFC.

We can summarize the mathematical assumptions and the mathematical features in the following table. Notice that, especially in the case of situation theory / theory of non-well-founded sets, it is not simple to make a clear decision concerning the following categories. A reason for this is the fact that different versions exist in the research community. Some have quite different features. Hence, our categorization is more tentative in character.[5]

[5] In the table below, GB systems denotes Gupta-Belnap systems and ST is a shorthand for situation theory.

	Kripke	GB Systems	ST
Theory	algebraic	algebraic	coalgebraic
Logic	non-classic	classic	classic
Syntax	classic	non-classic	non-classic
Semantics	classic	non-classic	classic
Definitions	classic	non-classic	classic
Set Theory	classic	classic	non-classic

The situation is a little bit more complicated as depicted in the table. Therefore, we need to discuss certain issues here.

Remark 16.1.3 (i) Notice that in a further development of situation theory, namely the so-called channel theory, even the underlying logic of the framework becomes non-classical.[6] The standard approach towards situation theory can be interpreted as an approach that preserves classical logic.

(ii) Whether the semantics of standard situation theory is classical or not cannot be trivially decided. The problem is that it is still an open problem whether there is a model for situation theory at all. A model can be defined if situation theory is appropriately restricted. For example, in [BarEt87], a semantics is developed. We did quite similar things in Sections 15.1 and 15.2. In the models of restricted fragments of situation theory, we have seen that ordinary model theory was used. Therefore, we choose to categorize the semantics of situation theory as classical.

(iii) We claim in the table above that Gupta-Belnap systems are non-classical in their syntax, because they allow definitions that count as syntactically ill-formed in classical accounts. Clearly the sentences that can be considered in the Gupta-Belnap systems are not syntactically non-classical, as for example in comparison with circular propositions in situation theory. Nevertheless, allowing these definitions yields new predicates that can only be defined in classical definition theory if one uses higher-order concepts.

(iv) Although revision theories are strong enough to define a non-well-founded universe[7], we cannot claim that revision theories do contain non-well-founded sets implicitly. It is only possible to define a universe allowing revision rules, relative to the given standard set theory ZFC.

(v) It should be noted that the categories *semantics* and *definitions* are not independent from each other. It is not possible to change the underlying definition theory without changing the semantics at the same time. This holds because if one allows circular definitions like the Gupta-Belnap account, a classical semantics cannot evaluate these definitions.

[6] Compare Subsection 16.2.3 for further information.

[7] Compare Subsection 9.3.1 for further information.

As an important insight, we can state that no account is purely classical in all respects. One or the other feature is non-classical in all frameworks. That is not extremely surprising if one takes into consideration that it is - according to Tarski's theorem - provably impossible to develop a theory of truth in the object language in strong enough languages.

We add some remarks concerning the different modifications in the classical theory. Clearly, the strongest modification of the mathematics used in order to develop theories for circularity is the modification of the underlying set theory. In logical applications, it is quite common to use different logics for different problems (logical pluralism). To use a different set theory is not as common. In this respect, the account which uses non-well-founded sets is the most radical framework. Concerning accounts that change the semantics and/or the definition theory, we can say that those accounts are at first sight not extremely different from other accounts. There is one additional feature in revision theories that must be taken into account: the strong systems $\mathbf{S}^*$ and $\mathbf{S}^{\#}$ are crucially defined using a sequence of hypotheses of length ORD. First, ordinals are used and second, the object that is constitutive for the definition of the system, namely the sequence of hypotheses, is no longer a set, but a proper class. We saw in Chapter 10 that it is possible to reduce the length of these sequences to sequences of countable length. We have to be aware of the fact that the techniques and mathematical assumptions used in revision theories are rather strong.

How strong is the assumption in situation theory that there are sets that are not contained in classical ZFC set theory? Changing the set theoretic universe is quite unfamiliar to most logicians. First, non-well-founded set theory is an extension of ZFC set theory and not a completely different set theory. The principal idea is simply to find a theory that includes all sets of ZFC but furthermore contains reflexive sets as well. This is a kind of extension that is well-known from other mathematical theories. Second, one can show that non-well-founded set theory is consistent relative to the consistency of ZFC set theory.[8] Hence, the theory of hypersets is no more problematic concerning consistency than ZFC. Third, it seems to be quite reasonable to assume that there are no a priori reasons why there should not be collections that can contain themselves. Axioms in set theory are usually seen as principles that are intuitively correct. In the case of the foundation axiom, it is not really clear why there should not be a set that contains itself. Considering these arguments, non-well-founded sets are not as strange as they probably appear to be at the very beginning. In total, we can say that - although using non-well-founded sets cannot be seen as a trivial extension of the set theoretic universe - it does not seem to be the case that there are a priori reasons against the usage of hypersets.

In the next subsection, we will mention some remarks concerning the philosophical assumptions and, in particular, a comparison of the philosophical assumptions of the different frameworks.

[8] Cf. [Ac88] and [BarMo96].

16.1.3 Philosophical Assumptions

We already considered some aspects of the philosophical assumptions of the different frameworks in Section 11.2. Here, we will add some more remarks concerning this topic.

We saw in Chapter 11 that Kripke's framework and the Gupta-Belnap systems are not different concerning the ontological status of the bearer of truth. In both theories, sentences of the given language L are the entities that can be true or false. In comparison with these two theories, situation theory differs. One of the central claims of situation theory was the statement that the bearers of truth are propositions, not sentences. Although the ontological status of propositions is less clear than the ontological status of sentences, the proponents of this choice argue that we need to specify in what situation one utters a certain sentence token. Contextual sensitivity becomes an important feature in situation theory as we saw in the difference between modeling the Liar sentence from a global perspective (where the Liar sentence is stably false) and a local perspective (where background information is included and the Liar sentence can be true or false).[9]

It is quite obvious that the account identifying propositions as bearers of truth is the more flexible one. Ignoring the contextual situation in which a sentence token is uttered can result in an approach that uses sentences as the bearer of truth. Sofar, Kripke's and Gupta's ideas are special cases of the more general account using propositions. The problem is that the introduction of a context is complicated and it is not a priori clear how this can be achieved in Kripke's and Gupta-Belnap's accounts. Clearly, the evaluation systems need to be adjusted to the new situation. Moreover, one has to clarify what it means to evaluate propositions in a revision sequence where every revision step changes the evaluation context. These points need to be addressed when someone tries to give a generalized account of revision theories.

The usage of urelements in the theory of hypersets is for many people quite uncommon. Most textbooks of set theory (as well as research papers) develop a set theory for pure sets. Important for the account of hypersets as examined in this work is the fact that urelements are crucially used in order to state a system of equations. In this respect, the question arises whether urelements are a necessary part of the theory or whether they are only used to simplify the representation. In fact, it is possible to get rid of the urelements. [Mo∞*b*] shows how to develop the theory of hypersets without urelements. Clearly, the abstract development in [Mo∞*b*] loses a lot of the intuitive plausibility of the development in Chapter 11. At the same time, the abstract framework of coalgebras generalizes the account and makes the large number of different types of systems superfluous. In this respect, the usage of urelements is not required.

From a philosophical perspective, urelements interpreted as atomic non-structured objects that can be used to create sets should not create ontological

[9] Compare Section 15.2 for further information.

problems. It is relatively undisputed that the attempt to speak about the real world forces us to introduce objects that are themselves non-sets. Only in this case one can speak about people in a discourse, about a bike, or an atomic power station, because all of these things are not pure sets.[10] That we use a pure class of these objects that are further specified in the particular application is motivated technically, because we ensure that we have always enough fresh urelements.

In general, the ontological status of sets is a strongly discussed topic in the philosophy of mathematics. Good introductions into different accounts of this discussion are [Ma90] for a realistic approach, [Fi80] for a nominalistic position, and [Ki83] for an empiristic approach. It is not possible to discuss the different motivations and ideas of these theories in this work. As a matter of fact, besides the classical problem of what the ontological status of sets are, we have the further problem of what the ontological status of urelements are. The easiest possibility to find a solution for this problem is to assume that urelements are physical or abstract objects in the real world. In this respect, materialism seems to be the natural interpretation of set theory with urelements.

A further distinction behind the ideas of the different frameworks can be detected: Whereas Kripke's account tries to remain in classical model theory and interprets all sentences with an extension that is given from the very beginning,[11] Gupta's account evaluates every sentence in revision sequences. The results of these evaluations determine the validity or the invalidity of particular sentences. In other words: whereas in fixed point theories the stage by stage process is hidden in the development of the theory (what is needed is simply the specification of a fixed point), Gupta incorporates the step-by-step process into the semantical system. The visibility of this process is a technical difference between the two theories. A judgment as to which of the two accounts is intuitively more appropriate is not trivial. Clearly, Gupta-Belnap systems are more transparent than Kripke's fixed point approach. Furthermore, it seems to be the case that reasoning about Liar-like sentences precisely presupposes the step-by-step evaluation of sentences. From this perspective, revision theories are closer to the intuition of what is going on when we reason about Liar-like sentences.

The next subsection summarizes some of the empirical facts. We already examined several examples of differences in the modeling of circularity in Chapter 10.

16.1.4 Empirical Facts

First, we restrict our attention to the problem of truth. We mentioned already that Kripke's account cannot model certain discourses where an information

[10] Clearly, even this is based on the assumption that not everything in the world is a pure set.

[11] Clearly, some sentences do not have extensions at all.

theoretic context is included. Examples for this lacking expressive strength are the Gupta puzzle as was discussed in length in Chapter 10 (together with the modified version of the Gupta puzzle). Other examples are all forms of logical tautologies.[12] In Kripke's framework, no logical tautology is preserved because of the fact that neither the three-valued logic nor the four-valued alternative logic has tautologies. Hence, it is impossible to represent tautologies. Furthermore, for Kripke's framework, the strengthened Liar is a problem and cannot be appropriately modeled in his theory. The reason for this problem is the finiteness of the possible truth values. Intuitively, we can lift the classical problem to another level.

All of these problems do not arise in Gupta-Belnap systems except for one problem. As we saw in Subsection 10.1.3, a slight modification of the Gupta puzzle is also a problem for Gupta-Belnap systems in the classical formulation like in [GuBe93]. Another problem for revision theories is the Truth-circle construction. Even here, one needs a modification of the theory itself in order to adjust the results of the modeling as was described in Subsection 10.1.2. Clearly, even the approach of Gupta and Belnap does not solve every problem. Without disqualifying Kripke's ideas, it is obvious that from an empirical point of view, revision theories are a better framework for circularity. They show better modelings of the phenomena, in the sense that they seem to be the more appropriate frameworks.

An important point is that in revision theories, there is no spelled out theory to represent other forms of circularity besides non-well-founded sentences of different forms. Circular phenomena occurring in theories of knowledge representation or the problem to distinguish common ground from private knowledge, cannot be easily modeled in Gupta-Belnap systems. One of the major problems is that revision theories do not include a theory of propositions. Propositions are most important in theories of knowledge representation, because propositions are the ontological entities we are talking about. Although in [GuBe93] it is explicitly mentioned that revision theories are considered as a framework to represent applications other than non-well-founded sentences as well, these ideas are far away from being spelled out. Further work needs to be done in order to decide to what extent such a representation in the revision theoretic account can be realized.

Concerning situation theory we can model a variety of different phenomena. Clearly, situation theory based on non-well-founded set theory is the most flexible theory for modeling circularity. The classical pathological sentences in natural language can be adequately represented in situation theory. Furthermore, in situation theory, it is possible to represent different perspectives as, for example, the global perspective versus the narrow perspective (as was described in Subsection 15.3.1). The high flexibility of situation theory is the reason for the ability to analyze further circular phenomena as the modeling of common ground and private knowledge.[13] Here, the important insight is to use propositions as bearers of truth. In total, one can state that the situation

[12] Cf. Section 6.1 for further information concerning that property of Kripke's account.

[13] Compare Subsection 15.3.2 for further information.

theoretic account is a very flexible one for modeling circularity. Problematic in this framework is the lack of an appropriate model for the overall theory. Concerning axiomatizations of the frameworks, we can state that all examined theories do have axiomatizations, although sometimes only as a second-order version.

The following table summarizes the results we mentioned and examined so far. It is clear that this table is not considered to be complete in any sense. Only some of the most important differences between the frameworks are mentioned without going into all details. We restrict our attention to the most important properties.

	Kripke	Gupta-Belnap systems	Situation Theory
Truth circle	*	not in standard theory	*
Strengthened Liar		*	*
Tautologies		*	*
Gupta puzzle		*	(*)
Modified G. puzzle		not in standard theory	(*)
CG vs. PK			*
Theory of propositions			*
Existence of Model	*	*	
Ex. of Axiomatization	*	*	*

Remark 16.1.4 (i) In the above table, CG vs. PK represent the difference between common ground and private knowledge.

(ii) Clearly, even in situation theory there are different possibilities of formulating the theory. One can examine weaker and stronger versions of situation theory. For example, the Gupta puzzle cannot be represented in standard situation theory developed in [BarEt87], because complex infons are needed and the concept of truth in a model needs to be adjusted for this example. That is the reason why we used brackets in the corresponding columns. In principle, the situation theoretic account is very flexible, but one has to develop refinements and modifications in order to apply it to particular phenomena.

(iii) We need to mention a further point. Clearly, there is the possibility extending revision theories to theories that can deal with propositions. Nowhere in the literature was an endeavor to develop such an extension. Hence, it seems to be the case that an introduction of contextual facts would require modifications in formulation of the semantical systems: there is no longer a fixed model that can count as the basis of every evaluation. These potential modifications need to be spelled out.

(iv) The problem of finding a model for situation theory is a severe problem for situation theory. Unfortunately, all attempts to find such a model failed. Although the lack of a model of the overall theory is not an final argument against the possibility of applications of situation theory, for a theory of circularity it seems to be necessary to require such a model. Here are desiderata that need to be fixed.

As a matter of fact, it can easily be seen that the considered theories are quite different concerning their empirical properties and their behavior in applications. The most flexible theory is obviously situation theory. Revision theories were developed in order to provide a theory of truth and not to develop a general account of circularity.[14] So, it is not surprising that they do not include a theory of propositions and knowledge representation.

In the last section, we mention some possible directions for further research and alternative accounts for a theory of circularity. We think that no end has been reached in the endeavor to model circularity. The next section is the attempt to give some ideas of possible future research.

16.2 Open Ends

Although we introduced the most important and most discussed frameworks for circularity, it is clear that there are

(a) other theories as well and
(b) there is a deeper context in which one can summarize the accounts.

We will give more explanations with respect to (b) here. First, we will formulate open questions concerning the developed theories in this work. Second, we will summarize some basic ideas of the so-called channel theory, an information theoretic account in order to model quite different kinds of reasoning. Third, we will discuss whether there are alternatives of the discussed frameworks. Finally, we need to ask again the question that was already formulated in Chapter 2: What is circularity?

16.2.1 Open Problems

We mention some open problems in this subsection that are connected with the frameworks developed so far. We do not discuss possible ways to reach a solution of the considered phenomena. This subsection attempts to give a list of open problems.

[14] The last Chapter in [GuBe93] can be read that revision theories should also provide a framework that can be applied to many different kinds of circularity. Even if this is correct, the theory was primarily developed in order to give an account for a theory of truth.

We begin with a problem that is connected with Part II of this work. The described problem has to do with the question of finding fixed points of operators that are defined on partially ordered sets.

Problem 16.2.1 *Assume $\Gamma : D \longrightarrow D$ is a functor defined on a CCPO $\langle D, \leq \rangle$. Give necessary and sufficient conditions for Γ, such that Γ has fixed points.*

There already exists a characterization for the question described in Problem 16.2.1, proven by the author in an unpublished manuscript. Unfortunately, this characterization is not very informative for applications we have in mind. The characterization mentioned can be given in terms of sub-CCPOs of the given CCPO $\langle D, \leq \rangle$. But this is not general enough in order to decide whether a given operator has fixed points or not. A very similar result was proven by Jose Martinez Fernandez using the theory of clones of functions.[15] It is not clear to the author whether the result of Martinez guarantees a compact representation of sufficient and necessary conditions of an operator in order to have fixed points. Recently, Martinez Fernandez is working on an extension of his result to predicate logic.

A natural question arises in the context of revision theories: Is it possible to develop a coalgebraic treatment of the revision theoretic account? Here, the problem is to introduce ordinal numbers to the theory of coalgebras in order to define the length of revision sequences.

Problem 16.2.2 *Can the semantical systems $\mathbf{S}^*$ and $\mathbf{S}^\#$ precisely be represented in a coalgebraic framework?*

If the length of revision sequences is ignored, it is relatively easy to reformulate revision theories in coalgebraic terms. But to specify an ordinal number α for an α-reflexive hypothesis cannot be achieved without explicitly introducing ordinal numbers.

A possible extension of revision theories is the introduction of second-order quantifications. This means that in a (possibly circular) definition it is not only allowed to quantify over individual variables but also over relation and function variables. What relations can be defined in such a theory? Furthermore, which complexity does the resulting system have?

Problem 16.2.3 *Determine the properties of revision theoretic systems in which second order quantifications are allowed.*

If we take into account that the semantical systems $\mathbf{S}^*$ and $\mathbf{S}^\#$ have complexity Π^1_2 although only first-order quantifications are allowed in circular definitions, it seems to be likely that in second-order versions the complexity of the resulting systems should be higher than the complexity of the original definitions.

[15] Cf. [Ma99].

Another question concerning revision theories concerns alternative semantical systems in which an evaluation of revision sequences can be achieved. In particular, it would be interesting if one could find a semantical system that has less complexity than $\mathbf{S}^{\#}$ and $\mathbf{S}^{*}$, but furthermore shows appropriate properties with respect to the modeling of the truth predicate.

Problem 16.2.4 *Find alternative semantical systems to the systems* $\mathbf{S}^{\#}$*,* $\mathbf{S}^{*}$*, and* $\mathbf{S}_n$*.*

We come to the theory of non-well-founded sets. We saw in Chapter 15 that the theory of hypersets can be developed completely in a coalgebraic framework. Although the theory of coalgebras is well-examined to a certain extent, from an abstract perspective, there is no complexity theory of coalgebras. Intuitively, what needs to be done is a characterization of *unfolded structures* of coalgebraically defined structures. We already saw in Fact 14.1.1 how we can unfold systems of equations in order to apply the flat anti-foundation axiom. For arbitrary structures something quite similar to the Flattening Lemma is needed, in order to develop a theory of complexity. The following Problem describes this precisely.

Problem 16.2.5 *Develop a complexity theory for the theory of coalgebras.*

An appropriate complexity theory of coalgebras should classify different collections of coalgebraic structures into complexity classes. The easiest way to achieve this would probably be first, a transformation of the coalgebraic structures into infinite algebraic structures and second, a classification of these unfolded structures.

At this point, we end our list of open problems concerning the considered formal frameworks for circularity. Clearly, it is not intended to give a complete list of open problems concerning the examined frameworks. Only some of the most important ones are mentioned. The next subsection deals with possible alternatives to the frameworks presented in this work.

16.2.2 Are there Alternatives?

In this subsection, we consider alternatives to the examined frameworks. Clearly, there are alternatives if one takes the enormous amount of works and papers into account that deals directly or indirectly with circularity. If one checks the literature concerning paradoxes in natural language one can find a variety of alternative accounts.[16] We think that the frameworks described in this work are the most promising ones. Additionally, they give an idea of the features of the underlying theories (concerning logic, set theory, semantics, syntax etc.) that can be modified in order to get an appropriate theory. One of the central claims of this work is that an appropriate theory cannot be classical

[16] Compare [Ya93, Ya85, Si93], but also from a computer science perspective the book [Tu90], only to mention some of them.

in all respects. In a certain sense that is more or less already a consequence of Tarski's theorem.

Clearly, there are alternatives. Consider the table in Subsection 16.1.2. One can change all non-empty subsets of parameters in order to find a possible framework for circularity. Take the five parameters logic, syntax, semantics, definition theory, and set theory. Then, we have $2^5 - 1 = 31$ possibilities to change possible parameters. And these are only the general ideas. Some theories may look quite different in their formulation although they may change precisely the same parameters. Hence, there are many possible alternatives.

What can be said about such alternatives? It is problematic to say something definitely about a potential theory that is not developed yet. Clearly, from an abstract point of view, all the alternatives must count as equal. From a more applied point of view, it is clear that there are strong intuitions behind each particular theory. For example, Kripke's account is inspired by the idea that there are sentences that are neither true nor false (namely the pathological ones). Situation theory is motivated by the idea that one needs a representation of propositions that contain themselves. And revision theories are inspired by the idea that circular definitions should explicitly be allowed. Therefore, for all these accounts there are relatively strong intuitions for or against a particular change in the classical theory. Hence, not all possible alternatives are equally justified from a practical perspective.

Because of the fact that we want to model real-world phenomena, one cannot ignore the motivating intuitions behind the formal representation. Therefore, it seems to be important to give at least some ideas concerning the different possibilities for modifications in a formal representation. The famous proponent for a change in the syntax in order to model circular sentences was Tarski.[17] In Tarski's account, pathological sentences are merely blocked in the object language, not modeled in a practically reliable way. The examined frameworks in this work are examples in which logical, semantical, definitional, and/or set theoretical parameters are changed. Clearly, the most fundamental change is the modification of the parameter set theory, because the mathematical objects themselves are modified. Logical adjustments are not as fundamental, because the mathematical entities we are speaking about remain the same. Additionally, logical pluralism and the usage of many different logics is quite common in many applications besides theories for circularity. Changing the semantics (hence the model theory) of a theory is on the other hand a relatively strong modification, because the usage of non-classical model theory is not very common. That is probably one of the reasons why revision theories are intuitively hard to handle if one considers the technical side of the relatively clear general idea.

We can conclude that there are technically a lot of possible alternatives. The intuitive plausibility and motivation of the described frameworks are definitely an advantage of the theories we considered in Parts II, III, and IV. Apart from technical details and parameters one can establish in the formulation of a

[17] In fact, Tarski tried to give a consistent theory of truth.

theory, the presented theories can count as the most important ones.

In the next subsection, we will consider (on a very informal level) channel theory as a possible further framework. We cannot introduce channel theory in length, because that would go far beyond the scope of this work.

16.2.3 Channel Theory

The paradigmatic theory for a flexible framework in order to represent various kinds of reasoning is the so-called channel theory.[18] Channel theory was developed in order to formulate a general theory of information transfer. Although the motivation for channel theory goes back to problems in situation theory and situation theory struggled with the modeling of circular propositions quite early in its history, circularity did not play any role in the development of that theory. The original motivation was to model constrains in natural language, like the constraint that kissing someone implies in one or the other way to touch someone (to mention one of the most prominent examples in [BarPe83]). Because classical logic is not an appropriate framework for these constraints, the developers of channel theory were looking for a more liberal theory.

We will not develop the theory in this subsection. That would go far behind the limits of what such a work can achieve. The attempt is to give an idea of the basic concepts used in channel theory and to specify the theory's main purpose. The basic idea on which channel theory is based is the type-token distinction. In channel theory, every token (of the real world or of the abstract world) is of a certain type. In a certain sense, the world can only be described if a certain conceptualization is already given. The interesting feature concerning channel theory is the fact that a dynamics in the change of certain conceptualizations is implemented on a very elementary level. The concept of an infomorphism guarantees that a dynamic change from a given categorization of tokens to another one can easily be described. We specify this in the following definition.

Definition 16.2.6 *(i) A classification* $\mathbf{A} = \langle tok(\mathbf{A}), typ(\mathbf{A}), \models_{\mathbf{A}} \rangle$ *is a triple, such that* $tok(\mathbf{A})$ *and* $typ(\mathbf{A})$ *are sets and* $\models \subseteq tok(\mathbf{A}) \times typ(\mathbf{A})$*. We call* $tok(\mathbf{A})$ *the tokens of* $\mathbf{A}$ *and* $typ(\mathbf{A})$ *the types of* $\mathbf{A}$*. The expression* $a \models_{\mathbf{A}} \alpha$ *denotes that* a *is of type* α*.*

(ii) Assume two classifications $\mathbf{A}$ *and* $\mathbf{B}$ *are given. An infomorphism* $f = \langle f^{\wedge}, f^{\vee} \rangle$ *is a pair of total functions, such that (a) and (b) holds:*

(a) $f^{\wedge} : typ(\mathbf{A}) \longrightarrow typ(\mathbf{B})$ *and* $f^{\vee} : tok(\mathbf{B}) \longrightarrow tok(\mathbf{A})$*.*
(b) $\forall b \in tok(\mathbf{B}) \forall \alpha \in typ(\mathbf{A}) : b \models_{\mathbf{B}} f^{\wedge}(\alpha) \longleftrightarrow f^{\vee}(b) \models_{\mathbf{A}} \alpha$*.*

The important idea of Definition 16.2.6(ii) is that infomorphisms offer a possibility to move from one classification **A** to another classification **B** (and vice versa). For example, if we have two classifications that describe one and the same phenomenon in two different ways (relative to given information theoretic

[18] [BarSel97] is the standard reference for this theory.

background), then we can move from the information coded in one classification to the information coded in the other classification.

How can the implemented dynamics of channel theory be used in order to represent circularity? We only give a rough idea of how this could be achieved. Notice that classifications can be interpreted as concepts that represent the information theoretic background. For example, a certain proposition β can be of type $True$. Given two different classifications **A** and **B** it should be possible - although this is not spelled out - to represent the dynamics of the reasoning about Liar-like propositions, namely the switching of the truth values that are assigned to the propositions. A modeling where we prefer not to evaluate Liar-like propositions is also possible. In this case, certain tokens β are neither of type $True$ nor of type $False$.

What can be said about other phenomena of circularity? At least a broad number of such phenomena could be modeled using channel theory. The problem is more or less the precise specification of the modeling. To spell out a certain application in terms of channel theory precisely, can be a rather complicated matter.[19] The reason for this is the explosion of notational complexity, if one introduces more concepts in channel theory, as for example channels, distributed systems, theories, and (local) logics. Clearly, the complexity of notation increases quite often when one tries to develop better and more appropriate models than alternative theories are able to offer. Therefore, it seems that this is the price we have to pay for further frameworks of circularity.

The described ideas are rather simple. Channel theory should be able to represent more than this. Because we are free to define information channels and logics on classifications, we can develop a variety of additional constraints on top of the dynamics of the classifications. For example, one could think of establishing a constraint that guarantees that logical tautologies are preserved. Notice that this does not necessarily block the representation of the flip-flop behavior of Liar-like sentences, because in channel theory, one can model a multi-level approach. In a certain sense, it should be possible to represent the global perspective (every Liar-like sentence is false and logical tautologies are true) as well as the local perspective (Liar-like sentences show a constantly non-stable behavior) in one framework.

The intuitively plausible idea to introduce the dynamics on a very elementary level seems to be a reason to make an attempt in this direction for future research. Much should be said about possible problems concerning notational complexity. Clearly, a solution for this problem is non-trivial. We will not deepen this idea here but rather leave space for further research in this area.

Our last subsection deals with the problem from the beginning: What is circularity? After studying different frameworks for circularity, we discuss whether we can reach a certain approximation of this problem.

[19] Compare examples in [BarSel97] to get a flavor of the involved complexity.

16.2.4 What is Circularity?

The more philosophical question remains to be reconsidered: *What is circularity*? In Section 2.6, we gave an approximation of the concept of circularity in terms of non-well-foundedness. We claimed that an entity is circular if and only if there are aspects of this entity that are non-well-founded. Furthermore, our conjecture was that a circular entity is pathological if it does not allow an appropriate well-founded representation. Notice that a concept like self-referentiality plays no role in these claims.[20] In order to explain circularity, we go back to a relatively clear notion. The definition using non-well-foundedness has a further advantage. A non-well-founded structure can be considered as mathematically well-defined and intuitively clear, whereas self-referentiality is mathematically less clear.

Does the above definition give us a tool in order to get a clearer picture of a structured discussion about circularity? We claim that this is in fact true. The reason for this is not only that we have an appropriate conceptualization at hand, but moreover that we have an appropriate mathematical model for non-well-founded phenomena as well. Notice that the paradigmatic mathematical model behind a non-well-founded definition is a coinductive definition, namely a definition that is category theoretically the dual of an inductive definition. Not the step-by-step creation of an object is the idea behind a coinductive definition, but merely a destruction (or breakdown) of a structure in its parts, where circles are explicitly allowed. To have a mathematical model at hand is a great advantage, because one can get a precise analysis of a phenomenon and a possibility to test the analysis in applications of the theory.

A remark should be added concerning the considered frameworks. Clearly, Kripke's approach as well as the Gupta-Belnap systems are based on classical algebraic theories and do not contain coalgebraic elements. Nevertheless, they provide appropriate theories for certain circular phenomena. In particular, Gupta-Belnap systems give a good approach towards a theory of pathological expressions in natural languages. Is this a counterexample of our explanation? I think it is not. The reason for this lies simply in the basic mathematical assumptions. Clearly, we work in classical set theory (ZFC) without hypersets. Hence, every coalgebraic account can in principle be represented using classical mathematics. In other words, every account for modeling circularity can be represented in an algebraic framework. The only restriction is that it can become necessary to consider infinite (because unfolded) objects.[21] This suffices to argue that the algebraic version of revision theories can be developed using pure algebraic tools and methods.

Is the problem of the Liar solved? Clearly, it is not solved in the sense that we have for every possible occurrence of circular phenomena an appropriate

[20] Quite often circularity is explained in terms of self-referentiality. According to our explanations one can get rid of this additional concept.

[21] In the semantical systems $\mathbf{S}^*$ and $\mathbf{S}^{\#}$, one is forced to consider infinite objects of length ORD.

answer. For every particular occurrence one has to adjust the framework. But the above question includes a problem. From a naive point of view, it seems difficult to determine what is meant by a solution of Liar-like paradoxes. Clearly, a paradox remains a paradox and cannot be made consistent by a mathematical modeling.[22] Hence, a solution of the Liar paradox would mean finding a framework that can represent paradoxical expressions without becoming paradoxical itself. The examined frameworks provide such theories and they are (at least relatively to the consistency of ZFC) consistent.

[22] Sometimes one has the feeling that precisely this is the attempt of certain authors.

Bibliography

[Ac78] P. **Aczel**, Frege structures and the notions of proposition, truth and set, in: The Kleene Symposium, edited by J. Barwise, H.J. Keisler, K. Kunen, Amsterdam 1978, pp.31-59.

[Ac88] P. **Aczel**, Non-well-founded sets, Cambridge 1988, CSLI Lecture Notes.

[Ac90] P. **Aczel**, Replacement Systems and the Axiomatization of Situation Theory, in: Situation Theory and Its Applications, edited by R. Cooper, K. Mukai, and J. Perry, Stanford 1990, pp.3-31.

[Ac96] P. **Aczel**, Generalized Set Theory, in: Logic, Language and Computation, edited by J. Seligman and D. Westerstahl, Stanford 1996, pp.1-17.

[AcLu91] P. **Aczel** and R. **Lunnon**, Universes and Parameters, in: Situation Theory and Its Applications, vol. 2, edited by J. Barwise et. al. Stanford 1991, pp.3-24.

[AcMe89] P. **Aczel** and N. **Mendler**, A final coalgebra theorem, in: Proceedings category theory and computer science, edited by D. Pitt, D. Ryeheard, P. Dybjer, A. Pitts, and A. Poigne, Lecture Notes in Computer Sciences, 1989, pp.357-365.

[An94a] A. **Antonelli**, The Complexity of Revision, in: Notre Dame Journal of Formal Logic, vol. 35, No.1, 1994, pp.67-72.

[An94b] A. **Antonelli**, Non-well-Founded Sets via Revision Rules, in: Journal of Philosophical Logic, vol.23, No.6, 1994, pp.633-680.

[An94c] A. **Antonelli**, A Revision-Theoretic Analysis of the Arithmetical Hierarchy, in: Notre Dame Journal of Formal Logic, vol.35, No.2, 1994, pp.204-218.

[An98] A. **Antonelli**, The Complexity of Revision, revised. Unpublished Manuscript, May 1998.

[ArAv96] O. **Arieli** and A. **Avron**, Reasoning with Logical Bilattices, in: Journal of Logic, Language and Information, vol. 5, 1996, pp.25-63.

[Au50] J. **Austin**, Truth, in: Proceedings of the Aristotelian Society, supp. Vol. XXIV, 1950, pp.110-128.

[Ba81] H. **Barendregt**, The λ-calculus, Amsterdam, North-Holland, 1981.

[BaMoSo∞] A. **Baltag**, L. **Moss** and S. **Solecki**, The Logic of Public Announcements and Common Knowledge, unpublished manuscript, available in the www under the address: `http://www.math.indiana.edu/home/moss/articles.html`.

[BaWe90] M. **Barr** and C. **Wells**, Category Theory for Computing Science, Prentice Hall, 1990.

[BaSu87] S. **Bartlett** and P. **Suber** eds., Self-Reference: Reflections on Reflexivity, Dordrecht, Nijhoff, 1987.

[Bar75] J. **Barwise**, Admissible Sets on Abstract Structures. An Approach to Definability Theory, Berlin 1975.

[Bar77] Handbook of Mathematical Logic, editors: J. **Barwise** with the cooperation of H. **Keisler**, Amsterdam, North-Holland 1977 (Studies in Logic and the Foundation of Mathematics, 90).

[Bar81] J. **Barwise**, Scenes and Other Situations, in: Journal of Philosophy, vol.59, 1981, pp.369-396.

[Bar90] J. **Barwise**, On the Model Theory of Common Knowledge, in: J. Barwise: The Situation in Logic, Cambridge, 1990 (CSLI Lecture Notes, No.17).

[Bar97] J. **Barwise**, Information and Impossibilities, in: Notre Dame Journal of Formal Logic, vol.38, No.4, 1997, pp.488-515.

[BarPe83] J. **Barwise** and J. **Perry**, Sittuations and Attitudes, MIT Press, Bradford Book 1983.

[BarEt87] J. **Barwise** and J. **Etchemendy**, The Liar. An Essay on Truth and Circularity, Cambridge 1987, CSLI Lecture Notes.

[BarEt90] J. **Barwise** and J. **Etchemendy**, Information, Infons, and Inference, in: Situation Theory and its Applications, edited by R. Cooper et. al., Stanford 1990, pp.33-78.

[BarMo96] J. **Barwise** and L. **Moss**, Vicious Circles. On the Mathematics of Non-Wellfounded Phenomena, Cambridge 1996, CSLI Lecture Notes.

[BarSel97] J. **Barwise** and J. **Seligman**, Information Flow. The Logic of Distributed Systems, Oxford 1997.

[Be82] N. **Belnap**, Gupta's rule of revision theory of truth, in: Journal of Philosophical Logic, vol.11, pp.103-116.

[Be78] J. **van Benthem**, Modal Corresponding Theory, Doctoral Dissertation, Mathematical Institute, University of Amsterdam 1978.

[Be83] J. **van Benthem**, The Logic of Time: a model-theoretic investigation into the varieties of temporal ontology and temporal discourse, Dordrecht, Reidel 1983.

[Be96] J. **van Benthem**, Exploring logical dynamics, Stanford, CSLI Publications, 1996.

[BeDo83] J. **van Benthem** and K. **Doets**, Higher-Order Logic, in: Handbook of Philosophical Logic, vol.1: Elements of Classical Logic, edited by D Gabbay and F. Guenthner, Reidel Dordrecht 1983, pp.275-329.

[Bo65] J. **Bochenski**, Formale Logik, Freiburg, München 1965.

[Bo97] L. **Bonjour**, In Defense of Pure Reason: A Rationalist Account of a Priori Justification, Cambridge 1997.

[Bre92] E. **Brendel**, Die Wahrheit über den Lügner. Eine philosophisch-logische Analyse der Antinomie des Lügners, Berlin, New York 1992.

[Br92] A. **Brueckner**, Semantic Answers to Skepticism, in: Pacific Philosophical Quarterly, vol.73, 1992, pp.200-219.

[Bu82] J. **Buridan**, John Buridan on self-reference: Chapter 8 of Buridan's Sophismata, translated with an introduction and a philosophical commentary by G. Hughes, Cambridge University Press 1982.

[Bu79] J. **Burges**, Axioms for Tense Logic I, in: Notre Dame Journal of Formal Logic, vol.23, 1979, pp.367-374.

[CaDa91] D. **Cain** and Z. **Damnjanovic**, On the weak Kleene Scheme in Kripke's Theory of Truth, in: Journal of Symbolic Logic, vol.56, no.4, 1991, pp.1452-1468.

[Ca86] J **Cargile**, Critical notice of Martin 1984, in: Mind, vol.95, pp.116-126, 1986.

[ChKe73] C. **Chang** and H. **Keisler**, Model Theory, Amsterdam, North-Holland, 1973.

[Ch93] A. **Chapuis**, Circularity, Truth, and the Liar Paradox, Doctoral Dissertation, Indiana University, Bloomington 1993.

[Ch96] A. **Chapuis**, Alternative Revision Theories of Truth, in: Journal of Philosophical Logic, vol.25, 1996, pp.399-423.

[Ch73] C. **Chihara**, Ontology and the Vicious-Circle Principle, Ithaca, New York, Cornell University Press, 1973.

[Ch77] R. **Chisholm**, Theory of Knowledge, Englewood Cliffs, New Jersey, 1977.

[CMCS98] Electronic Notes in Theoretical Computer Science, CMCS 1998: First Workshop on Coalgebraic Methods in Computer Science, Vol. 11, guest editors: B. Jacobs et.al., available under the domain name: `http://www.math.tulane.edu/ENTC.html`.

[DaM097] N. **Danner** and L. **Moss**, On the Foundation of Corecursion, in: Logic Journal of the IGPL, vol.5, No.2, 1997, pp.231-257.

[DaPr90] B.A. **Davey** and H.A. **Priestley**, Introduction to Lattices and Order, Cambridge 1990.

[Da90] D. **Davidson**, The Structure and Content of Truth, in: Journal of Philosophy, vol.LXXXVII, no.6, 1990, pp.279-329.

[Da92] J. **Davoren**, A Lazy Logician's Guide to Linear Logic, Technical Report, Department of Mathematics, Monash University, 1992.

[De91] K. **Devlin**, Logic and Information, Cambridge, New York 1991.

[De93] K. **Devlin**, The Joy of Sets. Undergraduate Texts in Mathematics; Berlin Springer 1993.

[Dr81] F. **Dretske**, Knowledge and the Flow of Information, MIT Press, Cambridge, Mass., 1981.

[Du78] M. **Dummett**, Truth and other Enigmas, Cambridge 1978.

[Du85] M. **Dunn**, Relevance Logic and Entailment, Handbook of Philosophical Logic, vol.3, ed. D. Gabbay and F. Guenthner, Dordrecht: Reidel, pp.117-224.

[DuEp77] Modern uses of multiple-valued logic, editors: M. **Dunn** and G. **Epstein**, Reidel, Dordrecht 1977.

[EbFl95] H.-D. **Ebbinghaus** and J. **Flum**, Finite Model Theory, Springer-Verlag, Berlin-Heidelberg 1995.

[Eh61] A. **Ehrenfeucht**, An application of games to the completeness problem for formalized theories, Fundamenta Mathemaaticae **49**, pp.129-141, 1961.

[EiMa45] S. **Eilenberg** and S. **MacLane**, General theory of natural equivalences, in: Transactions of the American Mathematical Society, vol. 58, 1945, pp.231-294.

[Fe84] S. **Feferman**, Toward Useful Type-Free Theories I, in: Recent Essays on Truth and the Liar Paradox, edited by R.L. Martin, Oxford, New York 1984, pp.237-287.

[Fe90] T. **Fernando**, On the Logic of Situation Theory, in: Situation Theory and its Applications, vol.1, edited by R.Cooper et. al., Cambridge 1990, pp.97-116.

[Fi80] H. **Field**, Science without Numbers: A Defense of Nominalism, Princeton 1980.

[Fi75] P. **Finsler**, Aufsätze zur Mengenlehre, Darmstadt, Wissenschaftliche Buchgesellschaft, 1975.

[Fit74] F. **Fitch**, Elements of Combinatory Logic, New Haven 1974.

[Fi81] M **Fitting**, Fundamentals of Generalized Recursion Theory, Amsterdam, North-Holland 1981 (Studies in Logic and the Foundations of Mathematics, vol.105).

[Fi87] M. **Fitting**, Computability Theory and Logic Programming, Oxford, 1987.

[Fi88] M. **Fitting**, Logic Programming on a topological bilattice, in: Fundamenta Informaticae, vol. 11, 1988, pp.209-218.

[Fi89] M. **Fitting**, Bilattices and the Theory of Truth, in: Journal of Philosophical Logic, vol. 18, 1989, pp.225-256.

[Fi91] M. **Fitting**, Bilattices and the Semantics of logic programming, in: The Journal of Logic Programming, vol. 11, 1991, pp.91-116.

[Fi93] M. **Fitting**, The family of stable models, in: The Journal of Logic Programming, vol. 17, 1993, pp.197-225.

[Fi94] M. **Fitting**, Kleene's three valued logics and their children, in: Fundamenta Informaticae, vol. 20, 1994, pp.113-131.

[FoHo83] M. **Forti** and F. **Honsell**, Set Theory with Free Construction Principles, in: Annali Scuola Normale Supeiore di Pisa, Classe di Scienze, vol. 10, 1983, pp.493-522.

[Fr54] R. **Fraïssé**, Sur quelques classifications des systémes de relations (English summary), Université d'Alger, Publications Scientifiques, Série A, **1**,pp.35-182, 1954.

[GaSt53] D. **Gale** and F. **Stewart**, Infinite games of perfect information, Annals of Mathematical Studies 28, 1953, pp.245-266.

[GeGr97] J. **Gerbrandy** and W. **Groeneveld**, Reasoning about information change, in: Journal of Logic, Language, and Information, vol.6, 1997, pp.147-169.

[Gi86] M.L. **Ginsberg**, Multi-valued logic, in: Proceedings AAAI-86, Fifth National Conference on Artificial Intelligence, Morgan Kaufmann Publishers, 1986, pp.243-247.

[Gi88] M.L. **Ginsberg**, Multi-valued Logics: A Uniform Approach to Inference in Artificial Intelligence, in: Journal of Computational Intelligence, vol. 4, no. 3, 1988, pp.265-316.

[GiLaTa89] J.-Y. **Girard**, Y. **Lafont** and P. **Taylor**, Proofs and Types, Cambridge University Press 1989, Cambridge Tracts in Theoretical Computer Sciences, vol. 7.

[Go31] K. **Gödel**, Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I, in: Monatshefte für Mathematik und Physik, vol. 38, 1931, pp.173-198.

[GoHa86] G. **Goos** and J. **Hartmanis**, Category Theory and Computer Programming, Lecture Notes in Computer Science No.240, edited by G. Goos and J. Hartmanis, Springer-Verlag, Berlin 1986.

[Gr78] G. **Grätzer**, General Lattice Theory, Basel, Stuttgart 1978 (Lehrbücher und Monographien aus dem Gebiet der exakten Wissenschaften: Math. Reihe, vol.52).

[GrNe08] K. **Grelling** and L. **Nelson**, Bemerkungen zu den Paradoxien von Russell und Burali-Forti, in: Abhandlungen der Fries'schen Schule, neue Folge 2, 1908, pp.301-334.

[Gr66] P. **Grice**, Meaning, in: Philosophical Review, vol. 66, 1966, pp. 377-388.

[Gr90] J. **Grimshaw**, Argument Structure, Cambridge, MIT Press, 1990.

[Gr94] W. **Groeneveld**, Dynamic Semantics and Circular Propositions, in: Journal of Philosophical Logic, vol.23, 1994, pp.267-306.

[Gu82] A. **Gupta**, Truth and Paradox, in: Journal of Philosophical Logic 11, 1982, pp.1-60.

[Gu89] A. **Gupta**, Remarks on Definitions and the Concept of Truth, in: Proceedings of the Aristotelian Society, vol.89, 1989, pp.227-246.

[GuBe93] A. **Gupta** and N. **Belnap**, The Revision Theory of Truth, MIT Press, Cambridge, Mass., 1993.

[Ha91] L. **Haegeman**, Introduction to Government and Binding Theory, Cambridge, Mass, 1991.

[Ha96] F. **Hamm**, Nominalizations, events, and facts, Ms., University of Tübingen 1996.

[Ha77] G. **Harman**, Review of Linguistic Behavior by Jonathan Bennett, in: Language, vol.53, 1977, pp.417-427.

[Hei82] I. **Heim**, The Semantics of Definite and Indefinite Noun Phrases in English, PhD Thesis, University of Massachusetts, Amherst, distributed as *Arbeitspapiere 73*, SFB 99 Konstanz, 1988.

[He50] L. **Henkin**, Completeness in the Theory of Types, in: Journal of Symbolic Logic, vol.15, 1950, pp.81-91.

[He82a] H. **Herzberger**, Notes on Naive Semantics, in: Journal of Philosophical Logic 11, 1982, pp.61-102.

[He82b] H. **Herzberger**, Naive Semantics and the Liar Paradox, in: Journal of Philosophy, 79, 1982, pp.479-497.

[He71] A. **Heyting**, Intuitionism: an introduction, third rev. edition, Amsterdam 1971.

[Hi80] J. **Higginbotham**, Pronoun and bound Variables, in: Linguistic Inquiry, vol.11, 1980, pp.679-708.

[Ho93] W. **Hodges**, Model Theory, Cambridge University Press, 1993.

[Ho79] D. **Hofstadter**, Gödel, Escher, Bach, New York, 1979.

[HoUl79] J. **Hopcroft** and J. **Ullman**, Introduction to Automata Theory, Languages and Computation, Reading (Mass.) 1979.

[Im99] N. **Immerman**, Descriptive complexity, New York, Springer, 1999.

[Ja95a] B.P.F. **Jacobs**, Inheritence and cofree constructions, CWI Report CS-R9564, Amsterdam 1995.

[Ja95b] B.P.F. **Jacobs**, Objects and classes, coalgebraically, CWI Reports CS-R9536, Amsterdam 1995.

[Ja96a] B.P.F. **Jacobs**, Automata and behaviours in categories of processes, CWI Report CS-R9607, Amsterdam 1996.

[Ja96b] B.P.F. **Jacobs**, Coalgebraic specifications and models of deterministic hybrid systems, CWI Report CS-R9609, Amsterdam 1996.

[JaRu97] B.P.F. **Jacobs** and J.J.M.M. **Rutten**, A Tutorial on (Co)Algebras and (Co)Induction. Bulletin of EATCS Vol. 62, 1997, pp. 222-259.

[Je78] T. **Jech**, Set Theory, Academic Press, London 1978.

[Ka84] H. **Kamp**, A Theory of Truth and Semantic Representation, in: Truth, Interpretation and Information, edited by J. Groenendijk et. al., Foris, Dordrecht 1984, pp.1-41.

[KaRe93] H. **Kamp** and U. **Reyle**, From Discourse to Logic, Kluver 1993.

[Ke95] A. **Kechris**, Classical Descriptive Set Theory, Springer-Verlag, Berlin, New York, 1995.

[Kep00] S. **Kepser**, A Coalgebraic Modelling of Head-Driven Phrase Structure Grammar, Technical Report of the SFB 441, University of Tübingen, 2000, available under the domain name: `http://tcl.sfs.nphil.uni-tuebingen.de/∼kepser/papers/hpsgcoalgebra.ps.gz`.

[Ki94] P. **King**, Reconciling Austinian and Russellian Accounts of the Liar Paradox, in: Journal of Philosophical Logic, vol.23, No.5, 1994, pp.451-494.

[Ki92] R. **Kirkham**, Theories of Truth, Cambridge, 1992.

[Ki83] P. **Kitcher**, The nature of mathematical knowledge, New York 1983.

[Kl52] S. **Kleene**, Introduction to Metamathematics, Groningen, Amsterdam 1952.

[Kl55] S. **Kleene**, Hierarchies of number theoretic predicates, in: Bull. Amer. Math. Soc. 1955, pp.193-213.

[KlVe65] S. **Kleene** and R. **Vesley**, The Foundations of Intuitionistic Mathematics, especially in relation to recursive functions, Amsterdam, North- Holland, 1965.

[Kn28] B. **Knaster**, Un théorème sur les fonctions d'ensembles, Annales Soc. Pol. Math.; vol.6; pp. 133-134.

[KoMö99] H. **Kolb** and U. **Mönnich**, The Mathematics of Syntactic Structure. Trees and their Logics, Studies in Generative Grammar, vol.44, Berlin, New York 1999.

[KoPa81] D. **Kozen** and R. **Parikh**, An elementary proof of the completeness of PDL, in: Theoretical Computer Science, 1981, pp.113-118.

[Kr93] P. **Kremer**, The Gupta-Belnap Systems $\mathbf{S}^*$ and $\mathbf{S}^\#$ are not Axiomatisable, in: Notre Dame Journal of Formal Logic, vol.34, No.4, 1993, pp.583-596.

[Kr75] S. **Kripke**, Outline of a Theory of Truth, in: Recent Essays on Truth and the Liar Paradox, edited by R.L. Martin, Oxford, New York 1984, pp.53-81.

[Ku96a] K.-U. **Kühnberger**, Wahrheitsprädikate in formalen Sprachen und Fixpunkte in algebraischen Strukturen, unpublished Master thesis, University of Tübingen, 1996.

[Ku96b] K.-U. **Kühnberger**, Fixpunktkonstruktionen auf partiell geordneten Mengen, Abstract in the proceedings of the Deutsche Mathematiker-Vereinigung, Jahrestagung 1996, p.303.

[Ku99] K.-U. **Kühnberger**, Characterization Theorems of Classes of Interlaced Bilattices, Technical Report of the Forschergruppe 'Logik in der Philosophie', 1999.

[Ku∞a] K.-U. **Kühnberger**, Tree Automata and Coalgebras, in preparation.

[KuLo] K.-U. **Kühnberger** and B. **Löwe**, A Game Characterization of Revision Theoretically Definable Reals, in preparation.

[Ku83] K. **Kunen**, Set Theory. An Introduction to Independence Proofs, First reprint, North-Holland, Amsterdam 1983.

[Le74] K. **Lehrer**, Knowledge, Oxford, 1974.

[Le69] D. **Lewis**, Convention: A Philosophical Study, Harvard University Press, Cambridge Mass., 1969.

[LöWe∞] B. **Löwe** and P. **Welch**, Set-Theoretic Absoluteness and the Revision Theory of Truth, unpublished manuscript.

[Lu91] R. **Lunnon**, Many Sorted Universes, SRDs, and Injective Sums, in: Situation Theory and Its Applications, vol.2, edited by J. Barwise et. al., Stanford 1991, pp.51-79.

[Ma71] S. **MacLane**, Categories for the Working Mathematician, Springer-Verlag, Berlin 1971.

[Ma90] P. **Maddy**, Realism in Mathematics, Oxford 1990.

[ManArb86] E. **Manes** and M. **Arbib**, Algebraic Approaches to Program Semantics, Springer, New York 1986.

[MaSh76] Z. **Manna** and A. **Shamir**, The theoretical aspects of the optimal fixed-point, in: Siam Journal of Computing, vol.6, 1976, pp.414-426.

[Ma84] R.L. **Martin**, Recent Essays on Truth and the Liar Paradox, edited by R.L. Martin, Oxford, New York 1984, pp.53-81.

[MaWo75] R.L. **Martin** and P.W. **Woodruff**, On representing 'true-in-*L*' in *L*, in: Recent Essays on Truth and the Liar Paradox, edited by R.L. Martin, Oxford, New York 1984, pp.53-81.

[Ma99] J. **Martinez Fernandez**, Teoría de la verdad para lenguajes autorreferentes: una solución parcial al problema del punto fijo, Doctoral dissertation, University of Valencia 1999.

[Ma96] F. **Mau**, Formale Sprachen mit Wahrheitsprädikat, Doctoral Dissertation University of Osnabrück 1996.

[Mc91] V. **McGee**, Truth, Vagueness, and Paradox. An Essay on the Logic of Truth, Hackett Publishing Company, Indianapolis, 1991.

[MiTo91] R. **Millner** and M. **Tofte**, Co-induction in relational semantics, in: Theoretical Computer Sciences, vol 87, pp.209-220.

[Mi17] D. **Mirimanoff**, Les antinomies de Russell et de Burali-Forti et le probleme fondamental de la theorie des ensembles, in: L' enseignement mathematique vol.19, 1917.

[Mo82] G.H. **Moore**, Zermelo's Axiom of Choice. It's Origins, Development, and Influene; Springer, New York 1982.

[Mo74] Y. **Moschovakis**, Elementary Induction on Abstract Structures, North- Holland, Amsterdam, 1974.

[Mo80] Y. **Moschovakis**, Descriptive Set Theory, North-Holland, Amsterdam, 1980.

[Mo94] Y. **Moschovakis**, Set Theory Notes; Undergraduate Texts in Mathematics, Heidelberg Springer 1994.

[Mo∞a] L. **Moss**, Coalgebraic Logic, Preprint, Indiana University, Bloomington, to appear.

[Mo∞b] L. **Moss**, Parametric Corecursion, Preprint, Indiana University, Bloomington, to appear.

[MoSel96] L. **Moss** and J. **Seligman**, Situation Theory, in: Handbook of Logic and Language, edited by J. van Bentham und A. ter Meulen, Elsevier 1996, pp.239-309.

[Mu91] K. **Mukai**, CLP(AFA): Coinductive Semantics of Horn Clauses with Compact Constraints, in: Situation Theory and its Applications, vol.2, edited by J. Barwise et. al., Stanford 1991, pp.179-214.

[Mu89] R. **Muskens**, Meaning and Partiality, Diss. University of Amsterdam 1989.

[Ob93] A. **Oberschelp**, Rekursionstheorie, Mannheim, Leipzig, Wien, Zürich 1993.

[Od89] P. **Odifreddi**, Classical Recursion Theory, Amsterdam, New York 1989.

[Pa70] B. **Pareigis**, Categories and Functors, Academic Press, New York, London 1970.

[PaMeWa93] B. **Partee**, A. **ter Meulen**, and R. **Wall**, Mathematical Methods in Linguistics, Dordrecht 1993.

[PoSa94] C. **Pollard** and I. **Sag**, Head-Driven Phrase Structure Grammar, Chicago 1994.

[Pr79] G. **Priest**, The logic of paradox, Journal of Philosophical Logic, vol. 8, 1979, pp.219-241.

[Pr87] G. **Priest**, Unstable solutions to the Liar paradox, in: Self-Reference: Reflextions on Reflexivity, edited by S. Bartlett and p. Suber, Martinus Nijhoff, 1987.

[Pu90] L. **Puntel**, Grundlagen einer Theorie der Wahrheit, Berlin, New York 1990.

[Pu81] H. **Putnam**, Brains in a vat, in: Reason, Truth and History, Cambridge, 1981.

[Qu88] W. **Quine**, Word and Object, Cambridge, Mass., MIT Press, 1988.

[Re∞] G. **Restall**, Consequences. An Introduction to Substructural Logics, unpublished Draft.

[Rh88] R. **Rheinwald**, Semantische Paradoxien, Typentheorie und ideale Sprache: Studien zur Sprachphilosophie Bertrand Russells, Berlin 1988.

[Ro67] H. **Rogers**, Theory of Recursive Functions and Effective Computability, Cambridge, Massachusetts 1967.

[RuWh27] B. **Russell** and A. **Whitehead**, Principia Mathematica, vol. I-III, Second Edition, Cambridge 1927.

[Rut95] J. **Rutten**, A calculus of transition systems (towards universal coalgebra), CWI Report CS-R9503, 1995.

[Rut96] J. **Rutten**, Universal Coalgebra: A Theory of Systems, CWI Report CS-R9652, 1996.

[RuTu93] J. **Rutten** and D. **Turi**, On the Foundation of Final Systems: non-standard sets, metric spaces, partial orders, in: Proceedings of the REX Workshop on Semantics: Foundations and Applications, ed.: W. de. Bakker, W.-P. de Roever, and G. Rozenberg, LNCS, vol.66, pp.477-530.

[Ry86] D. **Rydeheard**, Adjunctions, in: Category Theory and Computer Programming, Lecture Notes in Computer Science No.240, edited by G. Goos and J. Hartmanis, Springer-Verlag, Berlin 1986, pp.51-57.

[Sa90] G. **Sacks**, Higher Recursion Theory, Springer, Berlin, Heidelberg, New York, London 1990.

[Sa95] R. **Sainsbury**, Paradoxes, second edition, Cambridge 1995.

[Sch72] S. **Schiffer**, Meaning, Oxford University Press, Oxford 1972.

[Sc96] A. **Schöter**, Evidential Bilattice Logic and Lexical Inference, in: Journal of Logic, Language, and Information, vol. 5, 1996, pp.65-105.

[ScDo93] P. **Schroeder-Heister** and K. **Dosen**, Substructural Logics, Oxford, 1993, Studies in logic and computation, vol.2.

[Sc70] H. **Schubert**, Kategorien I/II, Two volumes, Springer-Verlag, Berlin, Heidelberg, New York 1970.

[Sc72] H. **Schubert**, Categories, translated from the German by Eva Gray, Springer-Verlag, Berlin, Heidelberg, New York 1972.

[Sc75] D. **Scott**, Combinators and Classes, in: Proceedings of the symposium λ-calculus and computer science theory, edited by C. Böhm, Berlin 1975, pp.1-26.

[Se83] J. **Searle**, Intentionality: An Essay in the Philosophy of Mind, Cambridge University Press, 1983.

[Se94] **Sextus Empiricus**, Outlines of Skepticism, translated by Julia Annas and Jonathan Barnes, Cambridge University Press, 1994.

[Sh94] M. **Sheard**, A Guide to Truth Predicates in the modern Era, Journal of Symbolic Logic, vol. 59, 1994, pp.1032-1054.

[Si58] W. **Sierpinski**, Cardinal and Ordinal Numbers; Panstwowe wydawnnictwo naukowe 1958.

[Si93] K. **Simmons**, Universality and the Liar: An Essay on Truth and the Diagonal Argument, Cambridge University Press, 1993.

[Sm85] C. **Smorynski**, Self-reference and Modal Logic, New York, Springer, 1985.

[Sm92] R. **Smullyan**, Gödel's Incompleteness Theorem, Oxford 1992.

[So87] R. **Soare**, Recursively Enumerable Sets and Degrees. A Study of Computational Functions and computably Generated Sets, Springer, Berlin, Heidelberg, New York 1987.

[So94] E. **Sosa**, Philosophical Scepticism and Epistemic Circularity, in: Aristotelian Society Supplementary, vol.68, 1994, pp.263-290.

[St78] J.R. **Strooker**, Introduction to categories, homological algebra and sheaf cohomology, Cambridge University Press, 1978.

[Su60] P. **Suppes**, Axiomatic Set Theory, New York Dover Publications, 1960.

[Ta55] A. **Tarski**, A lattice-theoretic fixpoint theorem and its applications, in: Pacific Journal of Mathematics, vol. 5, 1955, pp.285-309.

[Ta56] A. **Tarski**, The Concept of Truth in formalized Languages, in: A. Tarski: Logic, Semantics, and Metamathematics, Oxford 1956, pp.152-278.

[Tr92] A. **Troelstra**, Lectures on Linear Logic, CSLI Publications, Cambridge 1992.

[TrDa88] A. **Troelstra** and D. **van Dalen**, Constructivism in Mathematics, Amsterdam, 1988.

[TrSc96] A. **Troelstra** and H. **Schwichtenberg**, Basic proof theory, Cambridge, 1996, Cambridge tracts in theoretical computer science, vol.43.

[Tu96] D. **Turi**, Functorial Operational Semantics and its Denotational Dual, Ph.D. thesis, CWI, Amsterdam 1996.

[TuPl97] D. **Turi** and G. **Plotkin**, Towards a Mathematical Operational Semantics, in: Proceedings of LICS '97, IEEE Press 1997, pp.280-291.

[TuRu97] D. **Turi** and J. **Rutten**, On the foundations of final coalgebra semantics: non-well-founded sets, partial orders, metric spaces, in: Mathematical Structures in Computer Sciences, vol.8, no.5 1998, pp.481-540.

[Tu90] R. **Turner**, Truth and Modality for Knowledge Representation, London 1990.

[Ve96] K. **Vermeulen**, Merging without Mystery, in: Journal of Philosophical Logic, vol.25, 1996.

[Vic89] S. **Vickers**, Topology via Logic, Cambridge 1989, Cambridge tracts in theoretical computer science, vol.5.

[Vi84] A. **Visser**, Four-valued Semantics and the Liar, in: Journal of Philosophical Logic, vol. 13, 1984, pp.181-212.

[Vi89] A. **Visser**, Semantics and the Liar Paradox, in: Handbook of Philosophical Logic, vol. IV, edited by D. Gabbay and F. Guenthner, Reidel, Dordrecht 1989, pp.617-706.

[ViVe96] A. **Visser** und K. **Vermeulen**, Dynamic Bracketing and Discourse Representation, in: Notre Dame Journal of Formal Logic, 1996, vol.3.

[Wa99] T.A. **Warfield**, A Priori Knowledge of the World: Knowing the World by Knowing Our Brains, in: Skepticism. A Contemporary Reader, herausgegeben von K. DeRose und T.A. Warfield, New York, Oxford 1999, pp.76-90.

[Wel∞a] P. **Welch**, Eventually Infinite Time Turing Machine Degrees: Infinite Time Decidable Reals, *to appear in :* Journal of Symbolic Logic.

[Wel∞c] P. **Welch**, On Gupta-Belnap Revision Theories of Truth, in preparation.

[Ya85] S. **Yablo**, Truth and Reflection, Journal of Philosophical Logic, vol.14, 1985, pp.297-349.

[Ya93] A. **Yaqūb**, The Liar Speaks the Truth. A Defense of the Revision Theory of Truth, Oxford 1993.

Zusammenfassung

Diese Arbeit ist in vier Teile gegliedert. Der erste Teil ist eine generelle Einführung in die Themenproblematik. Dieser gliedert sich in ein Kapitel, das die in der Arbeit benutzte Notation klärt, in ein Kapitel, das die paradigmatischen Beispiele zirkulärer Phänomene diskutiert und ein Kapitel, das einen Überblick über die drei verbleibende Hauptteile der Arbeit gibt. Ein wichtiger Punkt in diesem ersten Teil ist der Versuch einer begrifflichen Klärung von Zirkularität, insbesondere im Verhältnis zur Nicht-Fundiertheit eines Phänomens. Diese Klärung stellt den philosophischen Kern der primär mathematisch-formalen Arbeit dar. Im zweiten Teil dieser Dissertation wird Kripke's Fixpunktkonstruktion für partiell definierte Wahrheitsprädikate diskutiert. Hierzu werden die algebraischen Grundlagen eingeführt und Probleme dieser Konstruktion aufgezeigt. Die zentralen Ergebnisse dieses zweiten Teils sind drei Charakterisierungstheoreme von Teilklassen überlappender Biverbände und deren Anwendungen. Im dritten Teil der Dissertation werden Revisionstheorien diskutiert und deren Adäquatheit bezüglich der Repräsentation von Wahrheitsprädikaten und Zirkularität im allgemeinen diskutiert. Hierbei kommen Untersuchungen zur Komplexität dieser Theorien, eine Einordnung der Revisionstheorien in einen breiteren thematischen Kontext, sowie deren empirische Eigenschaften ein besonderes Gewicht zu. Im letzten Teil der Arbeit wird Zirkularität auf der Ebene der Mengentheorie eingeführt und in Anwendungen bezüglich ihrer Tragfähigkeit überprüft. Die zentrale Leitidee in diesem Teil ist das Konzept einer coalgebraischen Modellierung. Insbesondere die situationstheoretische Modellierung von Wahrheit und die Repräsentation des Unterschieds zwischen privatem Wissen und allgemeinem Wissen wird in das Zentrum des Interesses gerückt. Ein Vergleich der untersuchten Ansätze wird im letzten Kapitel dieser Arbeit durchgeführt.