Paradox, Arithmetic and Nontransitive Logic

Inaugural-Dissertation
zur Erlangung des Doktorgrades der Philosophie
der Ludwig–Maximilians–Universität
München

vorgelegt von
Jonathan Georg Dittrich
aus
Karlsruhe
2020

Referent: Hannes Leitgeb
Koreferent: David Ripley
Tag der mündlichen Prüfung: 6.07.2020

Table of Contents

Acknowledgements

This dissertation is the result of a long journey including many discussions and many revisions. When I arrived at the LMU, I was still naive about the technical preliminaries that would be necessary to complete my work. It is due to the great people at the MCMP that I was able to close these gaps and complete my project.

I met many great people during the time of writing my dissertation and so I apologise for anyone who is missing here that had a fruitful impact on my work. First and foremost, I would like to thank Hannes Leitgeb, Lavinia Picollo and Norbert Gratzl for many great discussions, suggestions and overall help during my time working on this project. Furthermore, I would like to thank David Ripley, Thomas Schindler, Ole Hjortland and Volker Halbach for great discussions especially during conferences. A special thank you goes to Peter Schroeder-Heister and Luca Tranchini. Their discussion of Tennant's original paper on the non-normalisability of paradoxes in natural deduction during their seminar in Tübingen in 2014 sparked my interest in nontransitive theories of truth. Last but not least, I would like to thank the Diaphora project for the financial support.

Work on this dissertation has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement no. 675415.

Zusammenfassung

Diese Dissertation befasst sich mit formalen Wahrheitstheorien, die unter einer nicht transitiven Logik geschlossen sind. Formale Wahrheitstheorien sind typischerweise Erweiterungen einer arithmetischen Theorie, in unserem Fall Peano Arithmetik, durch Axiome oder Regeln, die das Verhalten eines einstelligen Wahrheitsprädikats betreffen. Ohne jegliche Einschränkungen sind solche Theorien inkonsistent und trivial auf Grund von Paradoxien. Solche Inkonsistenzen lassen sich dadurch ausräumen, dass man entweder die zugrunde liegende Logik oder die Regeln, die das Wahrheitsprädikat betreffen, einschränkt. Diese Dissertation befasst sich mit der ersten Option. Sie diskutiert die Möglichkeit und Vorteile, die Transitivität der Logik einzuschränken und damit eine Wahrheitstheorie auf Grundlage einer nicht transitiven Logik zu definieren.

Das erste Kapitel führt die nötigen technischen Begrifflichkeiten für das Verständnis der Materie ein. Dazu gehört die Definition der formalen Sprache der Arithmetik erweitert durch ein Wahrheitsprädikat, die Gödel-Kodierung, sowie ein passendes Sequenzen-Kalkül. Letzteres erlaubt es uns, die Verwendung der Annahme, dass die zugrundeliegende Folgerungsrelation transitiv ist, explizit zu machen. Dies geschieht durch die Verwendung einer Regel namens Cut. Zudem wird aufgezeigt, wie sich die gängigen Paradoxien im Zusammenhang mit dem Wahrheitsprädikat als Instanzen des arithmetischen Diagonalisierungslemma verstehen lassen und wie sich dieses im Sequenzenkalkül beweisen lässt. Darauf folgt eine Übersicht über Zweifel an der Transitivität aus der Literatur. Einzelne Ansätze werden erläutert und diskutiert, wobei wir zum Schluss kommen, dass die meisten Ansätze Probleme aufweisen weil sie Instanzen der Transitivität als problematisch deklarieren, die für mathematische Theorien notwendig sind.

Im folgenden Kapitel wird deswegen eine alternative Motivation für die Einschränkung der Transitivität im Sinne der Cut Regel erläutert. Es wird gezeigt, dass die Cut Regel äquivalent zu einer Regel ist, die die Konsistenz der bisherigen Teilbeweise annimmt. Eine solche Annahme kann jedoch in einer Theorie mit Paradoxien, die zu Inkonsistenzen führen, nicht gerechtfertigt sein. Vielmehr sollte die Cut Regel nur dann anwendbar sein, wenn man einen Grund für die Annahme der Konsistenz der bisherigen Teilbeweise hat. Eine Möglichkeit aus der Literatur, solche Einschränkungen zu formalisieren, besteht in der Verwendung einer Default Logik. Es wird jedoch aufgezeigt, dass die Mechanismen dieser Logiken aus Gründen der Komplexität der formalen Wahrheitstheorien nicht anwendbar sind.

Das vierte Kapitel entwickelt deswegen einen alternativen Ansatz zur Formalisierung der Konsistenzannahme. Cut wir dahingehend eingeschränkt, dass beide Prämissen in einer konsistenten Wahrheitstheorie bewiesen sein müssen. Durch die Verwendung einer konsistenten Theorie wird die Konsistenz der bisherigen Teilbeweise garantiert. Es wird gezeigt, dass die resultierende Theorie in vielerlei Hinsicht fruchtbar ist: Es können sowohl alle Theoreme der klassischen Logik bewiesen werden, sowie alle Instanzen des T-Schemas und darüber hinaus auch alle Theoreme der stärksten klassischen Wahrheitstheorie.

Im fünften Kapitel wird ein systematischer Ansatz entwickelt, Wahrheitstheorien zu bewerten und zu vergleichen. Dafür werden verschiedene Werte sowohl aus der Literatur der Wahrheitstheorien als auch der Wissenschaftsphilosophie herangezogen. Zudem werden unterschiedliche Rollen des Wahrheitsprädikats, die substitutionelle und die quantifizierende, unterschieden. Neben den Werten ist eine Wahrheitstheorie auch vor allem daran zu messen, inwieweit das Wahrheitsprädikat in einer solchen Theorie diese Rollen erfüllen kann. Es wird argumentiert, dass der hier vorgestellte nicht transitive Ansatz sowohl in Hinblick auf die Rollen des Wahrheitsprädikats, als auch in Hinblick auf die relevanten Werte anderen Wahrheitstheorien der Literatur überlegen ist.

Das letzte Kapitel diskutiert Einwände gegen nichtklassische Wahrheitstheorien, zu dem auch unser nicht transitiver Ansatz zählt. Dabei wird auf Einwände eingegangen, die gegen die Revision klassischer Logik im Allgemeinen und solche, die gegen die Revision der Transitivität im Speziellen argumentieren. Die Konklusion gibt letztlich einen Überblick über die neuen Erkenntnisse und Verbesserungen im Bereich der formalen Wahrheitstheorien im Vergleich zur bestehenden Literatur.

Chapter 1

Introduction

This dissertation is concerned with motivating, developing and defending nontransitive theories of truth over Peano Arithmetic (PA). Its main goal is to show that such a nontransitive theory of truth is the only theory capable of maintaining all functional roles of the truth predicate: the substitutional and the quantificational roles. By the substitutional roles we mean that the theory ought to prove ϕ iff it proves $T\ulcorner\phi\urcorner$ (potentially even if ϕ is a subformula of a complex formula such as a conditional) and that it proves all instances of the T-schema $T\ulcorner\phi\urcorner \leftrightarrow \phi$. A theory fulfils the quantificational role if its axioms governing the truth-predicate are strong enough to mimick as much second-order quantification as possible. Where the literature on classical theories of truth has focused primarily on the fulfilment of the quantificational role, the nonclassical literature is very much obsessed with the substitutional roles.

The problem of having a theory of truth fulfilling both the substitutional and quantificational (or already just the full substitutional) role are paradoxes of truth such as the Liar. Where the Liar is a sentence which informally says about itself that it is not true, we can show that it is both true and not true, which typically allows us to conclude any formula whatsoever. This problem is overcome in the current approach by blocking the use of transitivity principles under certain conditions.

Chapter two is concerned with the technical preliminaries necessary for our investigations. We begin by defining our formal language of truth and define a typical Gödel-coding scheme in order to encode expressions of the language into numerals. As structural aspects of the consequence relation such as transitivity play an important role, we will study formal theories of truth as sequent calculi consisting of rules for the logical constants, arithmetic, the truth predicate as well as the structural rules. The property of transitivity will be expressed by the rule of Cut. Last, we give an overview over formalised versions of the most prominent truth-theoretic paradoxes in the previously defined sequent calculus on the basis of the weak and strong diagonal lemma.

The third chapter explores different reasons why transitivity has been rejected or restricted so far. This includes reasons from relevance logic, proof-theoretic semantics and bilateralism, an inferentialist theory of meaning. We argue that these motivations suffer from the fact that they dismiss Cut in domains where the rule is necessary and well justified. Last, we discuss proposals of the literature on how to restrict Cut as a rule. Similar to before, we argue that the proposals so far are not apt to recover as much Cut as necessary for different purposes.

Building on the failure of motivations to restrict Cut properly in the previous chapter, the fourth chapter introduces a new motivation of this kind. We show that Cut turns out to be a rule, which implicitly assumes consistency. This is achieved by independently motivating a rule which clearly makes such an implicit consistency assumption and then proving that it is equivalent to Cut. This understanding of Cut then suggests to restrict it to cases in which this assumption is true. This is to say that Cut is applicable iff the subderivations of the premise sequents do not contain an inconsistency. We discuss how this idea can be applied to a theory of truth formulated as a sequent calculus and dismiss traditional ideas of capturing a consistency constraint in terms of a default condition due to issues of complexity. Rather, we show how this consistency constraint can be easily and effectively be captured by demanding that the Cut-premises must be derived in some theory of truth known to be consistent.

The fifth chapter is dedicated to the study of nontransitive theories of truth which can be obtained from this restriction. We start with the most severe version of this restriction by blocking all applications of Cut where the premises are not derived in PAT (i.e. the theory of PA formulated in the language containing the truth predicate). Already this weak theory, called NTT, has interesting properties as it preserves all classically valid inferences for the full language with the truth predicate and thus it preserves everything which is provable in PAT. This initial theory is strengthened by embedding strong classical theories of truth into it. Given a classical theory of truth S, we construct a nontransitive theory of transparent truth NTT[S] in which Cut is only blocked if at least one of its premises were derived using an instance of a rule not contained in S. We study the case of the classical theory $\mathsf{UTB}(\mathsf{Z}_2)$ which is particularly strong with respect to the quantificational role of the truth predicate as it proves all translated sequents of full second-order arithmetic Z_2. Different formulations including elimination rules or indices are explored. Last, we show that Cut is the only rule which can be restricted this way to obtain a nontrivial theory of truth.

The following two chapters motivate and defend the developed nontransitive approach. Chapter six gives a systematic overview over desiderata for theories of truth as they have been formulated in the literature. We isolate a set of virtues for theories of truth, which aim to preserve the full functionality of the truth predicate and show that our nontransitive approach fulfils these desiderata. Further, we discuss theoretical virtues which sometimes pop up in the literature on theories of truth as well. We approach this issue from a systematic overview over theoretical virtues in the natural sciences.

Chapter seven presents various objections against revising classical logic in general or Cut in particular as well as against the approach of blocking Cut in order to escape other paradoxes such as those of validity. We successfully defend our nontransitive approach against these objections by highlighting that it does not share most issues and weaknesses of well known nonclassical alternatives. The final chapter concludes by highlighting the advancements of the dissertation in applying nontransitive logics over the current literature.

Chapter 2

Technical Preliminaries

This chapter is concerned with the technical preliminaries necessary for our investigation into paradoxes and their structure. We begin with some basics of Arithmetic and Gödel-coding, which are used as a syntactic background for the truth-predicate and which allow us to generate a wide range of paradoxical sentences. We further introduce an appropriate sequent calculus to formalise paradoxical arguments while making assumptions about transitivity in them explicit. The chapter closes with formal proofs of the most prominent truth-theoretic paradoxes.

2.1 Syntax

We are concerned with first-order formal theories of truth based on Peano Arithmetic (PA). First, we need to define the languages in which we will work. $\mathcal{L}_{PA}$, the language of Peano Arithmetic is defined using the following vocabulary:

$$\neg \mid \; \rightarrow \; \mid \forall \mid \; = \; \mid \; \times \; \mid \; + \; \mid S \mid \overline{0} \mid$$

It also contains a stack of variables $x_1, x_2, ...$ and function symbols $\underset{.}{f}_1, \underset{.}{f}_2, ...$ for all primitive recursive functions. Where f is a function, we denote its representing function symbol in our formal language by $\underset{.}{f}$. S denotes the successor function and $\overline{0}$ the number 0, $\times$ and $+$ represent multiplication and addition respectively.

The language of truth, $\mathcal{L}_T$, is defined by adding to $\mathcal{L}_{PA}$ a unary truth-predicate T. Terms and formulae of our languages are defined inductively as usual.

If a variable in ϕ is within the scope of a quantifier for the same value, it is called bound, otherwise it is a free variable. If a formula contains no free variables, it is a sentence. Where n is a natural number, we write $\overline{n}$ for its numeral. For example, the intended interpretation of the numeral $\overline{0}$ is 0. If $\overline{n}$ is a numeral, then $S\overline{n}$ is its successor numeral.

2.1.1 Coding

Informally, a truth-predicate T should apply to sentences (or propositions), not terms. For example, we might want to say '2 + 2 = 4' is true or 'Snow is white' is true. However, our formal definition above only allows applications of T on terms made up of $\overline{0}$ and function symbols but no sentences. To overcome this syntactic problem, we interpret some terms as names of sentences of our language. The standard way to do this is by Gödel-coding. We follow (Smith, 2013: p.136) here in setting up a coding scheme.

The goal of such a coding procedure is to give to each expression of $\mathcal{L}_T$ a term which encodes it, s.t. i) the term is itself contained in $\mathcal{L}_T$ and ii) there is exactly one code for each expression of $\mathcal{L}_T$. To do this, we first set up basic codes by associating each symbol of the language (except variables) with an odd number. For example, $\neg$ is associated with the basic code 1, $\rightarrow$ with 3, $\forall$ with 5 etc. Variables are then associated with even numbers. Having set up the basic codes, we can construct codes for complex expressions based on the basic codes of their constituents.

Let $\pi_0, \pi_1, ...$ be the succession of prime numbers and $c_1, c_2, ..., c_n$ the basic codes. An expression e of $\mathcal{L}_T$ can be thought of as the ordered list $s_1, ..., s_n$ of symbols s of $\mathcal{L}_T$. Where c_n is the basic code of s_n, the Gödel code of e is defined as

$$\pi_0^{c_1} \times \pi_1^{c_2} \times ... \times \pi_k^{c_n}$$

Since, by the Fundamental Theorem of Arithmetic, every natural number can be rewritten as a unique factorisation into primes, it is guaranteed that every expression

e of the language has at most one code. Of course there are enough primes for all expressions of the language, thus there is exactly one coding expression. Where ϕ is a formula of $\mathcal{L}_T$, we write $\#\phi$ for its Gödel code.

Note, however, that $\#\phi$ is not an expression of $\mathcal{L}_T$ itself, since it is a natural number! However, we can use the corresponding numeral, which is contained in $\mathcal{L}_T$. Where $\#\phi$ is the Gödel code of a formula, we will denote the numeral of that Gödel-code by $\ulcorner \phi \urcorner$. Thus whenever we write $T\ulcorner \phi \urcorner$, this is just a tidy way of writing an expression of the form $TSSS...\overline{0}$.

We denote a primitive recursive substitution function by $\underset{\cdot}{s}$. Let ϕ be a formula with exactly x free. Then $\underset{\cdot}{s}(\ulcorner \phi \urcorner, t)$ is to be read as the numeral of the code of the result of replacing all occurrences of x in ϕ by t. We abbreviate $\underset{\cdot}{s}(\ulcorner \phi \urcorner, t)$ by $\ulcorner \phi \underset{\cdot}{t} \urcorner$. This allows us to express quantification 'into' the truth-predicate by writing e.g. $\forall x T\ulcorner \phi \underset{\cdot}{x} \urcorner$.

2.2 Sequent Calculi

As we investigate the nature of paradoxes, we need to set up a precise formal framework in which we can study them. Since we are particularly interested in the *structure* of paradoxes, we ought to choose a framework which makes this structure explicit. One type of calculus, which is predominantly used for this purpose is the sequent calculus introduced by Gentzen in 1934 (see (Gentzen, 1969) for the collected works).

Where many other calculi deal with premises and conclusions in the form of formulae, a sequent calculus works via manipulating sequents. A sequent is an expression of the form $\Gamma \Rightarrow \Delta$, where Γ and Δ are sets of formulae and $\Rightarrow$ is the sequent arrow. The sequent arrow expresses a relation of consequence between the sets. The left-hand side of the sequent arrow (in our case Γ) is interpreted conjunctively, whereas the right-hand side (here Δ) is read disjunctively. So $\Gamma = \{\phi_1, ..., \phi_n\}$ is read as $\phi_1 \wedge \phi_2 \wedge ... \wedge \phi_n$ and $\Delta = \{\psi_1, ..., \psi_n\}$ is understood as $\psi_1 \vee \psi_2 \vee ... \vee \psi_n$. To highlight some formula in a set, we write it separately next to the set, so $\phi, \Gamma \Rightarrow \Delta$ is an abbreviation of $\{\phi\} \cup \Gamma \Rightarrow \Delta$.

There are many different ways to interpret the consequence claim of the sequent arrow. For an overview see e.g. (Paoli, 2013). Right now, it is sufficient for our purposes to understand the sequent arrow as the claim that the left-hand side entails, proves or implies the right-hand side in some informal sense. We will later introduce a more sophisticated, inferentialist interpretation of sequents.

Sequents can then be manipulated and put in relation to each other by sequent rules (we will use sequent rules and rules interchangeably). Although a system S does not explicitly contain a particular rule R, it can be the case that S proves the conclusion of R given a proof of its premise(s) . Similarly, it sometimes happens that eliminating a rule from our system does not affect what is in principle provable within it. In these cases we speak of admissible or eliminable rules respectively:

Definition 2.2.1. Admissible rule

A rule R is said to be admissible in a system S iff: if S proves the conclusion sequent of R, then it proves its premise(s).

Definition 2.2.2. Eliminable rule

A rule R is said to be eliminable in a system S whenever S contains R but S without R proves the same sequents as S itself.

2.2.1 Rules For Logical Constants

We will consider the following, classical rules for the logical constants of $\mathcal{L}_T$:

$$\frac{\Gamma \Rightarrow \Delta, \phi}{\neg\phi, \Gamma \Rightarrow \Delta}\neg L \qquad \frac{\phi, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg\phi}\neg R$$

$$\frac{\Gamma \Rightarrow \Delta, \phi \qquad \psi, \Gamma \Rightarrow \Delta}{\phi \to \psi, \Gamma \Rightarrow \Delta}\to L \qquad \frac{\phi, \Gamma \Rightarrow \Delta, \psi}{\Gamma \Rightarrow \Delta, \phi \to \psi}\to R$$

$$\frac{\phi s, \Gamma \Rightarrow \Delta}{\forall x\phi x, \Gamma \Rightarrow \Delta}\forall L \qquad \frac{\Gamma \Rightarrow \Delta, \phi y}{\Gamma \Rightarrow \Delta, \forall x\phi x}\forall R$$

Where in $\forall R$, y must be an eigenvariable, i.e. it must not occur in the conclusion sequent. The formula(e) written explicitly in the premise of the rule is (are) called the active formula(e) of an application of a rule, the explicit formula(e) in the conclusion the principal formula(e). All other formulae are called side-formulae or the context. All rules concerned with logical constants are sometimes called operational rules.

Our language $\mathcal{L}_T$ also includes a sign for identity, for which we can give the following rules. First, it will make things easier to add initial sequents for all identities true and false:

$$\frac{}{\Gamma \Rightarrow \Delta, s = t} = T \qquad \frac{}{s = t, \Gamma \Rightarrow \Delta} = F$$

Where in $= T$, $s = t$ must hold and for $= F$ it must be the case that $s \neq t$. These rules are not purely logical since they allow for arithmetical instances such as $\Rightarrow \overline{0} = \overline{0}$. However, they are particularly handy for our purposes and the fact that they are not purely logical does not seem to bring about any problems. Second, we will also want to make use of substitutions based on identity claims. Those can be made via the following rules on either the left- or the right-hand side:

$$\frac{\phi s, \Gamma \Rightarrow \Delta}{s = t, \phi t, \Gamma \Rightarrow \Delta} SubL \qquad \frac{\Gamma \Rightarrow \Delta, \phi s}{s = t, \Gamma \Rightarrow \Delta, \phi t} SubR$$

Last, it will come in handy for many derivations to have a rule around which lets one eliminate true identities on the left hand side of the sequent:

$$\frac{s = t, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta} = E$$

where $s = t$ must be true and where s or t involve function symbols, the respective functions must be primitive recursive. Note that $= E$ makes $= T$ admissible by the rule of Reflexivity (see below). For we can derive any sequent of the form $s = t, \Gamma \Rightarrow \Delta, s = t$ and then conclude $\Gamma \Rightarrow \Delta, s = t$ by $= E$. Despite this redundancy, it is simply practical and saves time in derivations to have both rules around.

2.2.2 Rules For Structure

The following rules are not concerned with any kind of vocabulary, be it logical (as above) or arithmetical (as below). Rather, these rules express certain structural features of the consequence relation of our theory such as reflexivity, transitivity and monotonicity. Our working theory will only include two of these rules, the others will be shown to be admissible. The first structural rule included in our system is the rule of Reflexivity:

$$\frac{}{\phi, \Gamma \Rightarrow \Delta, \phi}\text{ Ref}$$

Alternatively, Reflexivity is sometimes formulated as $\phi \Rightarrow \phi$ without the arbitrary contexts Γ, Δ. A nice feature of using the formulation above is that it makes the rules of Weakening admissible (see 2.2.3). These are the following:

$$\frac{\Gamma \Rightarrow \Delta}{\phi, \Gamma \Rightarrow \Delta}\text{ WL} \qquad \frac{\Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \phi}\text{ WR}$$

Another set of structural rules is that of Contraction:

$$\frac{\phi, \phi, \Gamma \Rightarrow \Delta}{\phi, \Gamma \Rightarrow \Delta}\text{ CL} \qquad \frac{\Gamma \Rightarrow \Delta, \phi, \phi}{\Gamma \Rightarrow \Delta, \phi}\text{ CR}$$

Due to the way in which sequents are defined in terms of sets of formulae, CL and CR can easily be shown to be admissible (regardless of other rules):

Lemma 2.2.1. CL and CR are admissible in NT.

Proof. Sequents $\Gamma \Rightarrow \Delta$ are defined in terms of sets Γ, Δ of formulae. Since sets are insensitive to the number of occurrences of elements, the sequents $\phi, \phi, \Gamma \Rightarrow \Delta$ and $\phi, \Gamma \Rightarrow \Delta$ are identical. The same holds for $\Gamma \Rightarrow \Delta, \phi, \phi$ and $\Gamma \Rightarrow \Delta, \phi$. Thus, trivially, the conclusion of CL or CR is provable whenever its premise is provable. □

The final and most important structural rule we will discuss is the rule of Cut. This rule can be formulated in at least two ways, which turn out to be equivalent in our setup:

$$\frac{\Gamma \Rightarrow \Delta, \phi \qquad \phi, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}\, Cut \qquad \frac{\Gamma \Rightarrow \Delta, \phi \qquad \phi, \Gamma' \Rightarrow \Delta'}{\Gamma, \Gamma' \Rightarrow \Delta, \Delta'}\, Cut^*$$

The left formulation Cut is called context-sharing, the right one Cut^* is called context-free. The distinction is based on the side formulae of the Cut-premises. In the context-sharing version, both premises must contain the same side formulae, this requirement is lifted in the context-free version. It is easy to show that by admissible Weakening and Contraction (more on this below), the two formulations are equivalent, i.e. they prove the same sequents. Due to this equivalence, we will use the two formulations rather liberally. When we talk of Cut as a rule, we mean the context-sharing version. However, for simplicity and readability reasons, we sometimes apply Cut even though the contexts are not identical thereby implicitly using the context-free formulation.

In the rule of Cut, we call ϕ the Cut-formula. The rule makes explicit the transitivity of the consequence relation of our logic. Considering a simple example, an application of Cut allows us to infer from the premises $\phi \Rightarrow \psi$ and $\psi \Rightarrow \epsilon$ to $\phi \Rightarrow \epsilon$. In other words, if ϕ implies ψ and ψ implies ϵ, then ϕ implies ϵ. The beauty of the sequent calculus here is that transitivity is made into an explicit feature on the level of derivations. In other calculi such as natural deduction, the consequence relation is of course transitive as well but we can only talk about transitivity on the metalevel[1]. In sequent calculus, we can spot where transitivity comes into play in a derivation simply by looking at whether and where the derivation contains applications of Cut. In the case of paradoxes, this allows us to check, where they depend on an application of Cut and thus on the assumption that transitivity of the consequence relation holds. Once this structure is made explicit, we can find a way around paradoxical conclusions by suitable restrictions on Cut.

It can be shown for certain formulations of classical logic[2] with identity that

[1] In fact, there is an explicit and well-understood connection between sequent calculi and natural deduction calculi here. Applications of Cut translate into non-normal derivations in natural deduction and vice versa. These are derivations where a logical constant is first introduced and later eliminated by an elimination rule (see e.g. (von Plato, 2003; Prawitz, 1965)

[2] Gentzens *Hauptsatz* is quite sensitive to the way in which e.g. the operational rules are formulated. Instead of $\neg$L, we may add $\phi, \neg\phi, \Gamma \Rightarrow \Delta$ to our system which together with Cut makes $\neg$L admissible. However, Cut in this system without the $\neg$L rule is not eliminable (or admissible) as can easily be checked.

Cut is an eliminable (or admissible) rule. This is known as the Cut-elimination (Cut-admissibility) theorem or Gentzen's *Hauptsatz*.

2.2.3 Rules For Arithmetic

As we will see soon, typical paradoxes of truth can be formally reconstructed as instances of the weak or strong diagonal lemma provable in many arithmetical theories. So in order to investigate paradoxes in our calculus, it should be strong enough to prove these lemmata for $\mathcal{L}_T$. This is ensured by adding rules for the axioms of Robinson Arithmetic (Q) and for the axiom schema of Induction, obtaining a calculus for Peano Arithmetic (PA). One way to provide a sequent calculus for Q is to simply add its axioms as initial sequents[3] (see (Takeuti, 1987)):

$$\frac{}{0 = St, \Gamma \Rightarrow \Delta}\,\text{Q1} \qquad \frac{}{Sr = St, \Gamma \Rightarrow \Delta, r = t}\,\text{Q2}$$

$$\frac{}{\Gamma \Rightarrow \Delta, r + 0 = r}\,\text{Q3} \qquad \frac{}{\Gamma \Rightarrow \Delta, r + St = S(r + t)}\,\text{Q4}$$

$$\frac{}{\Gamma \Rightarrow \Delta, r \times 0 = 0}\,\text{Q5} \qquad \frac{}{\Gamma \Rightarrow \Delta, r \times St = (r \times t) + t}\,\text{Q6}$$

$$\frac{}{\forall x \neg (r = Sx), \Gamma \Rightarrow \Delta, r = 0}\,\text{Q7}$$

Whilst Cut-elimination holds for the system of classical logic with identity above, it no longer holds after adding these sequents for Q. However, there are ways around this issue. A formulation of Q in terms of rules rather than initial sequents was provided by Negri and von Plato in (Negri and Von Plato, 2011). The system is deductively equivalent to one in which we add the sequents above, yet Cut is an eliminable (admissible) rule. For the purpose of simplicity, we will stick to the formulation of Q above using initial sequents. Although we will later put restrictions on Cut, this will not affect what is derivable arithmetically.

In order to work with PA instead of just Q, we need to further add the axiom schema of Induction. Axiomatically, Induction is typically written as

---

[3] $Q7$ is formulated a bit differently with the universal quantified formula on the left instead of the negated universal on the right. The two formulations are deductively equivalent, this is simply done in order to maintain the invertibility of the right $\neg$-rule.

$$(\phi\bar{0} \wedge (\phi x \rightarrow \phi Sx)) \rightarrow \forall x \phi x$$

A straightforward formulation of the schema as a rule is the following:

$$\frac{\Gamma \Rightarrow \Delta, \phi\bar{0} \qquad \phi x, \Gamma \Rightarrow \Delta, \phi Sx}{\Gamma \Rightarrow \Delta, \forall x \phi x} \text{ Ind}$$

While it is possible to rewrite the rules for Q in order to regain Cut-elimination, this is not possible once Induction is around. At least not if we assume proofs composed of sequent rules to be finite constructions. A technical proof of this fact is given in (Troelstra and Schwichtenberg, 2000). One option to regain the eliminability of Cut in arithmetical theories is to replace Induction (and the right quantifier rule) by the infinitary ω-rule (see e.g. (Schütte, 2012) and (Buchholz, 1997) for a translation between the finitary and infinitary case). However, the ω-rule as an infinitary rule is only semi-formal. Without going into the details of the discussion about the rule, we will take this as sufficient reason to abandon the possibility for our purposes.

Call the collection of all of the above rules formulated in $\mathcal{L}_{PA}$ the theory PA and PAT the theory which is the result of formulating the rules in $\mathcal{L}_T$. Assuming that PA is consistent (which we do), PAT is of course consistent as well, since all we did was to add a predicate into the language. We did not add any further principles, axioms or rules which would allow for a derivation of an inconsistency. In order to study paradoxes, we need to strengthen PAT by adding rules for T.

2.2.4 Rules For Truth

In studying the truth-predicate T proof-theoretically, we treat it as a primitive symbol, which is only governed by the principles we add to our theory in the form of sequent rules. Tarski, prominently argued that ideally, a theory of truth should include all instances of what is now known as the T-schema (or T-(bi)conditionals):

$$\phi \leftrightarrow T\ulcorner \phi \urcorner$$

In order for all these conditionals to be provable in our theory of truth, we add the following introduction rules for T:

$$\frac{\phi, \Gamma \Rightarrow \Delta}{T\ulcorner\phi\urcorner, \Gamma \Rightarrow \Delta}\,\text{T1} \qquad \frac{\Gamma \Rightarrow \Delta, \phi}{\Gamma \Rightarrow \Delta, T\ulcorner\phi\urcorner}\,\text{T2}$$

That these rules are adequate to capture the instances of the T-schema in a sequent calculus setting can be shown by proving the conditionals:

$$\frac{\dfrac{\phi \Rightarrow \phi}{T\ulcorner\phi\urcorner \Rightarrow \phi}\,\text{T1}}{\Rightarrow T\ulcorner\phi\urcorner \to \phi}\,{\to}\text{R} \qquad \frac{\dfrac{\phi \Rightarrow \phi}{\phi \Rightarrow T\ulcorner\phi\urcorner}\,\text{T2}}{\Rightarrow \phi \to T\ulcorner\phi\urcorner}\,{\to}\text{R}$$

Besides these introduction rules for the truth predicate, we may also consider according elimination rules T3 and T4:

$$\frac{T\ulcorner\phi\urcorner, \Gamma \Rightarrow \Delta}{\phi, \Gamma \Rightarrow \Delta}\,\text{T3} \qquad \frac{\Gamma \Rightarrow \Delta, T\ulcorner\phi\urcorner}{\Gamma \Rightarrow \Delta, \phi}\,\text{T4}$$

However, we will later show that in the presence of our introduction rules T1 and T2, these rules are admissible and so we will not add them to our main theory of naive truth. These rules are strongly connected to an often desired property of transparency:

Definition 2.2.3. Transparent truth predicate

A theory of truth S is said to have a transparent truth predicate iff: $\vdash_S T\ulcorner\phi\urcorner, \Gamma \Rightarrow \Delta$ iff $\vdash_S \phi, \Gamma \Rightarrow \Delta$ and $\vdash_S \Gamma \Rightarrow \Delta, T\ulcorner\phi\urcorner$ iff $\vdash_S \Gamma \Rightarrow \Delta, \phi$.

Clearly, any theory closed under T1-T4 (regardless of whether the rule is included in the system or merely admissible) has a transparent truth predicate. Besides instances of the T-schema, axiomatic theories of truth often consider a quantified version of the biconditionals, the uniform T-schema:

$$\forall x_1, ..., x_n(T\ulcorner\phi(\dot{x}_1, ..., \dot{x}_n)\urcorner \leftrightarrow \phi(x_1, ..., x_n))$$

It is straightforward to derive all instances of the uniform T-schema using our substitution function $\underset{\cdot}{s}$:

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\phi y \Rightarrow \phi y}{\phi y \Rightarrow T\ulcorner \phi y \urcorner}\;\text{T2}}{\ulcorner \phi y \urcorner = \underset{\cdot}{s}(\ulcorner \phi \urcorner, y), \phi y \Rightarrow T\underset{\cdot}{s}(\ulcorner \phi \urcorner, y)}\;\text{SubR}}{\phi y \Rightarrow T\underset{\cdot}{s}(\ulcorner \phi \urcorner, y)}\;= E}{\Rightarrow \phi y \to T\underset{\cdot}{s}(\ulcorner \phi \urcorner, y)}\;\to}{\Rightarrow \forall x(\phi x \to T\underset{\cdot}{s}(\ulcorner \phi \urcorner, x))}\;\forall\text{R}}{\Rightarrow \forall x(\phi x \to T\ulcorner \phi \underset{\cdot}{x} \urcorner)}
$$

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\phi y \Rightarrow \phi y}{T\ulcorner \phi y \urcorner \Rightarrow \phi y}\;\text{T1}}{\ulcorner \phi y \urcorner = \underset{\cdot}{s}(\ulcorner \phi \urcorner, y), T\underset{\cdot}{s}(\ulcorner \phi \urcorner, y) \Rightarrow \phi y}\;\text{SubL}}{T\underset{\cdot}{s}(\ulcorner \phi \urcorner, y) \Rightarrow \phi y}\;= E}{\Rightarrow T\underset{\cdot}{s}(\ulcorner \phi \urcorner, y) \to \phi y}\;\to}{\Rightarrow \forall x(T\underset{\cdot}{s}(\ulcorner \phi \urcorner, x) \to \phi x)}\;\forall\text{R}}{\Rightarrow \forall x(T\ulcorner \phi \underset{\cdot}{x} \urcorner \to \phi x)}
$$

The last step in each of the derivations is simply the notational convention for the substitution function. Using more sophisticated substitution functions (see e.g. Cantini, 1990)then gives us the more general schema with multiple quantifiers as we introduced it above. Having set up our theory of naive truth NT, we proceed by showing some of its basic properties and defining the graph-theoretic structure of proofs in sequent calculi.

Last, we will consider some additional truth-theoretic principles. These are not concerned with the introduction or elimination of the truth-predicate. Rather, they impose that the connectives and the quantifier interact with the truth predicate in a compositional way. In Hilbert-style calculi, the compositional principles for $\neg, \to$ and $\forall$ are typically formulated by the following axioms:

$$\forall s(Sent(s) \to (\neg Ts \leftrightarrow T\underset{\cdot}{\neg} s))$$

$$\forall s, t(Sent(s) \wedge Sent(t) \to ((Ts \to Tt) \leftrightarrow T(s \underset{\cdot}{\to} t)))$$

$$\forall v \forall x(Sent(\underset{\cdot}{\forall} vx) \to (T(\underset{\cdot}{\forall} vx)) \leftrightarrow \forall t Ts[t/v])$$

Here *Sent* is a predicate applied to exactly those terms t, which denote a sentence. The new function symbols $\underset{\cdot}{\neg}, \underset{\cdot}{\rightarrow}, \underset{\cdot}{\forall}$ denote p.r. functions, which take the numeral of a code of (a) formula(e) and give back the numeral of the code of the formula obtained by applying the respective logical constant. So given a formula ϕ, $\underset{\cdot}{\neg}\ulcorner\phi\urcorner$ denotes $\ulcorner\neg\phi\urcorner$.

A standard translation of these axioms into sequent rules is given in (Halbach, 2011: p.71):

$$\frac{\Gamma, \neg Tt \Rightarrow \Delta}{\Gamma, Sent(t), T\underset{\cdot}{\neg}t \Rightarrow \Delta}\ \neg\mathrm{CL} \qquad \frac{\Gamma \Rightarrow \neg Tt, \Delta}{\Gamma, Sent(t) \Rightarrow T\underset{\cdot}{\neg}t, \Delta}\ \neg\mathrm{CR}$$

$$\frac{\Gamma, Tt \rightarrow Ts \Rightarrow \Delta}{\Gamma, Sent(t\underset{\cdot}{\rightarrow}s), T(t\underset{\cdot}{\rightarrow}s) \Rightarrow \Delta}\ \rightarrow\mathrm{CL} \qquad \frac{\Gamma \Rightarrow Tt \rightarrow Ts, \Delta}{\Gamma, Sent(t\underset{\cdot}{\rightarrow}s) \Rightarrow T(t\underset{\cdot}{\rightarrow}s), \Delta}\ \rightarrow\mathrm{CR}$$

$$\frac{\Gamma, \forall t Ts[t/v] \Rightarrow \Delta}{\Gamma, Sent(\underset{\cdot}{\forall}vs), T(\underset{\cdot}{\forall}vs) \Rightarrow \Delta}\ \forall\mathrm{CL} \qquad \frac{\Gamma \Rightarrow \forall t Ts[t/v], \Delta}{\Gamma, Sent(\underset{\cdot}{\forall}vs) \Rightarrow T(\underset{\cdot}{\forall}vs), \Delta}\ \forall\mathrm{CR}$$

Note that T1 and T2 together with the rules for logical constants are sufficient to prove all instances of compositional schemata such as $\neg T\ulcorner\phi\urcorner \leftrightarrow T\underset{\cdot}{\neg}\ulcorner\phi\urcorner$:

$$\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{\phi \Rightarrow \phi}{\Rightarrow \phi, \neg\phi}\ \neg\mathrm{R}}{\Rightarrow T\ulcorner\phi\urcorner, \neg\phi}\ \mathrm{T2}}{\Rightarrow T\ulcorner\phi\urcorner, T\ulcorner\neg\phi\urcorner}\ \mathrm{T2}}{\neg T\ulcorner\phi\urcorner \Rightarrow T\ulcorner\neg\phi\urcorner}\ \neg\mathrm{L}}{\ulcorner\neg\phi\urcorner = \underset{\cdot}{\neg}\ulcorner\phi\urcorner, \neg T\ulcorner\phi\urcorner \Rightarrow T\underset{\cdot}{\neg}\ulcorner\phi\urcorner}\ SubR}{\neg T\ulcorner\phi\urcorner \Rightarrow T\underset{\cdot}{\neg}\ulcorner\phi\urcorner}\ =E}{\Rightarrow \neg T\ulcorner\phi\urcorner \rightarrow T\underset{\cdot}{\neg}\ulcorner\phi\urcorner}\ \rightarrow\mathrm{R}$$

The other direction and schemata are provable in the analogous way. However, T1 and T2 are too weak to prove the universally quantified versions, expressed in the axioms above, in any consistent subtheory of PAT + T1 and T2 (which is why they are typically added as additional axioms in the first place). But given the compositional principles as the above sequent rules, it is easy to show that the universally quantified sentences expressing compositional truth are provable. The theory of naive truth NT is defined by extending PAT via T1, T2 (for all formulae ϕ) and the compositional principles.

2.2.5 Inversion And Admissibility

Applications of sequent rules can be put together to form a proof with the structure of a rooted tree. This structure is intuitive and we already used it above to prove sequents in our sequent calculus but for a formally precise definition see (Bimbó, 2014: p.356ff).

Being presented with a sequent in a proof of NT including a complex formula, it will come in handy to show that there is also a proof of the premise(s) of the rule introducing the relevant logical constant.

Definition 2.2.4. Invertibility of R

A rule R in a system S is said to be invertible in S iff: if S proves an instance of the conclusion sequent of R,then it proves the respective instance of the premise(s) of R.

To give an easy example, invertibility of ¬R means that if S proves $\Gamma \Rightarrow \Delta, \neg\phi$, then it also proves $\phi, \Gamma \Rightarrow \Delta$.

Theorem 2.2.2. Invertibility of →, ¬.

The rules ¬L, ¬R, →L and →R are invertible in NT.

Proof. We give a typical argument by induction on the height of the proof tree as depicted in (Negri et al., 2008). For the case of height 0, consider the rule ¬L – the other cases work analogously. If $\neg\phi, \Gamma \Rightarrow \Delta$ is an initial sequent, we distinguish the following cases. If $\Gamma \Rightarrow \Delta$ is an initial sequent, then so is $\Gamma \Rightarrow \Delta, \phi$. If $\Gamma \Rightarrow \Delta$ is itself not an initial sequent, this must be because $\neg\phi \in \Delta$. In this case we have $\phi, \Gamma \Rightarrow \Delta', \phi$ as an initial sequent (where $\Delta' = \Delta \setminus \{\neg\phi\}$). An application of ¬R gives us $\Gamma \Rightarrow \Delta, \phi$. For the induction step, assume invertibility for height n. This time we consider the case of →R, again the other cases work analogously. Assume a proof ending in $\Gamma \Rightarrow \Delta, \phi \to \psi$. If the last inference has the conditional as its principal formula, we are done. If not, apply the induction hypothesis to the previous line containing the conditional. The same strategy goes through for →L. □

It will come in handy later not to have the rules of Weakening, T3 and T4 in our theory of truth. Since we still want to make sure that our systems are closed under these rules, we prove them to be admissible. But proving the admissibility of any rule in NT does not tell us much, since the theory is trivial. Thus any rule would be admissible. Instead, we define a nontrivial subtheory of NT, ST, by dropping all instances of Cut. The name ST is also used by (Cobreros et al., 2012) where ST stands for Strict-Tolerant and refers to a non-trivial theory of truth closed under T1-T4, in which Cut is dropped completely.

The next two lemmata show the admissibility of the rules of Weakening as well as the elimination rules for T. We referred to these results earlier in the text and here the full proofs are finally given.

Lemma 2.2.3. WL and WR are admissible in ST.

Proof. The proof is by induction on the height of derivations. For the base case, if $\Gamma \Rightarrow \Delta$ is an initial sequent, so is $\phi, \Gamma \Rightarrow \Delta$ and $\Gamma \Rightarrow \Delta, \phi$. For the induction step, apply the induction hypothesis to the previous line(s) in the derivation. □

Lemma 2.2.4. T3 and T4 are admissible in ST.

$$\frac{T\ulcorner\phi\urcorner, \Gamma \Rightarrow \Delta}{\phi, \Gamma \Rightarrow \Delta}\,\text{T3} \qquad \frac{\Gamma \Rightarrow \Delta, T\ulcorner\phi\urcorner}{\Gamma \Rightarrow \Delta, \phi}\,\text{T4}$$

are admissible in ST.

Proof. The proof is by Induction on the height of the derivation. Here we only discuss the case of T3, the proof for T4 is completely analogous.

For the base case, if $T\ulcorner\phi\urcorner, \Gamma \Rightarrow \Delta$ is an initial sequent, we distinguish two cases: If $T\ulcorner\phi\urcorner \notin \Delta$, then $\phi, \Gamma \Rightarrow \Delta$ is an initial sequent as well. If $T\ulcorner\phi\urcorner \in \Delta$, then replace the initial sequent by the short proof

$$\frac{\phi, \Gamma \Rightarrow \Delta', \phi}{\phi, \Gamma \Rightarrow \Delta}\,\text{T2}$$

For the Induction step assume the Induction hypothesis that the admissibility holds for proofs of height n. If $T\ulcorner\phi\urcorner$ is not principal, then apply the Induction hypothesis to the previous line (in the case of multiple premises, apply the Induction hypothesis to any line containing $T\ulcorner\phi\urcorner$). If the T-formula is principal, it could have been introduced by either a T-rule or a substitution rule. In the case of the T-rule, the admissibility is immediate by taking the premise sequent of the T-inference. If the T-formula was introduced by a substitution rule, the derivation looks as follows:

$$\dfrac{\begin{array}{c}\vdots\\ Tt, \Gamma \Rightarrow \Delta\end{array}}{t = \ulcorner\phi\urcorner, T\ulcorner\phi\urcorner, \Gamma \Rightarrow \Delta}\text{ SubR}$$

Where t could be either a closed or an open term. Consider the case that t is a closed term. Either $t = \ulcorner\phi\urcorner$ does or does not hold. If it does not, then the desired conclusion sequent $t = \ulcorner\phi\urcorner, \phi, \Gamma \Rightarrow \Delta$ is an instance of =F. If $t = \ulcorner\phi\urcorner$ holds, we distinguish cases based on whether t is a numeral or not. If it is a numeral and since $t = \ulcorner\phi\urcorner$ holds, Tt and $T\ulcorner\phi\urcorner$ are actually the same formula. So the conclusion follows by applying the Induction Hypothesis to the previous line.

If t is not a numeral, the occurrence of Tt could have either been introduced by Reflexivity or by a substitution rule. In the case of Reflexivity, consider the respective initial sequent, which is of the form $Tt, \Gamma' \Rightarrow \Delta'$. If $Tt \notin \Delta'$, then replace the initial sequent by $\phi, \Gamma' \Rightarrow \Delta'$ and eliminate any rules below in which Tt (or conclusions derived from it) were active. If $Tt \in \Delta'$, then the initial sequent has the form $Tt, \Gamma' \Rightarrow \Delta^*, Tt$. In this case, replace the initial sequent by $\phi, \Gamma' \Rightarrow \Delta^*, \phi$ and plug in the following proof:

$$\begin{array}{c}\dfrac{\dfrac{\phi, \Gamma' \Rightarrow \Delta^*, \phi}{\phi, \Gamma' \Rightarrow \Delta^*, T\ulcorner\phi\urcorner}}{t = \ulcorner\phi\urcorner, \phi, \Gamma' \Rightarrow \Delta^*, Tt}\\ \Pi\\ t = \ulcorner\phi\urcorner, \phi, \Gamma \Rightarrow \Delta\end{array}$$

Where Π are the corresponding steps of the original proof. The last case to consider is in which Tt was introduced by a substitution rule (assuming that it is principal, otherwise apply the Induction Hypothesis to the previous line). If so, then the substitution rule introduced some identity $t = s$. If s is a closed term as well, then apply the corresponding step from above depending on whether or not $s = t$ holds.

If the T-formula was obtained by a substitution with an open term, consider how the active formula in this substitution was obtained. If it was introduced by an instance of Reflexivity, apply the same strategy as above for closed terms. If it was introduced by an instance of a substitution rule, the proof has the following form:

$$\dfrac{\dfrac{Tx, \Gamma' \Rightarrow \Delta}{x = t, Tt, \Gamma' \Rightarrow \Delta}}{t = \ulcorner \phi \urcorner, T\ulcorner \phi \urcorner, \Gamma \Rightarrow \Delta}$$

Where $\Gamma = \Gamma \cup \{x = t\}$. Again, go up the proof and find the formula in the trace of Tx, s.t. the active formula from which it was obtained involves either only closed terms or was obtained by an instance of Reflexivity. Apply the respective strategy from the case of closed terms above. □

2.3 Paradoxes

As Tarski's theorem (see Tarski, 1956) famously shows, no sufficiently strong classical theory such as NT can be closed under all instances of the T-schema, i.e. all instances of T1 and T2:

Theorem 2.3.1. Tarski's theorem.

Let S be a consistent, classical theory including a predicate T, strong enough to derive equivalences of the strong or weak diagonal lemma. Then S cannot prove all instances of $\phi \leftrightarrow T\ulcorner \phi \urcorner$.

Proof. The proof proceeds via a typical Liar-sentence argument. Consider the instance of the weak diagonal lemma $\lambda \leftrightarrow \neg T\ulcorner \lambda \urcorner$. Since S is classical, either $T\ulcorner \lambda \urcorner$ or $\neg T\ulcorner \lambda \urcorner$ must be the case. If $T\ulcorner \lambda \urcorner$ is the case, then by the T-schema λ follows. Using the equivalence from the weak diagonal lemma, we get $\neg T\ulcorner \lambda \urcorner$, a contradiction. Assuming $\neg T\ulcorner \lambda \urcorner$, using first the weak diagonal lemma and then the other direction of the T-schema gives us $T\ulcorner \lambda \urcorner$, which again contradicts our assumption. Contradiction. □

The problem with the naive theory of truth NT which figures in the proof of Tarski's theorem is the problem of paradoxes of truth. This section is dedicated to the study of such paradoxes starting from an informal perspective and moving on to formalisations of paradoxical arguments in sequent calculus.

Informally, paradoxes are often defined as *An argument with (prima facie) true premises, valid reasoning but an unacceptable conclusion* (see e.g. (Sainsbury, 2009)). NT formalises the relevant premises as rules and ensures classically valid reasoning. We distinguish some forms of unacceptable conclusions and fix a formal definition of a paradox in NT. The most prominent candidate of truth-theoretic paradoxes is the Liar paradox, which we already discussed in the proof of Tarski's theorem. The Liar entails a contradiction: In NT we can show that it is both true and not true. The reason for the unacceptability of this conclusion in classical logic shared by all authors is that a contradiction entails every formula of the language (by the so-called principle of explosion). But not all paradoxes prove an arbitrary formula of the system via a detour over a contradiction. A Curry-sentence for some arbitrary formula ϕ lets us prove ϕ directly.

The Curry-sentence says about itself that if it is true, then some arbitrary conclusion ϕ follows. Consider for example the conclusion that 'the moon is made of green cheese'. Then our particular Curry-sentence says about itself that if it is true, then the moon is made of green cheese. Again, classical logic dictates that the sentence must either be true or false. If it is true, then also what it says must be true. So it holds that if it is true, the moon is made of green cheese. By modus ponens it follows that the moon is made of green cheese. If it is not true, it also follows that if it is true, then the moon is made of green cheese (the conditional is trivially true, since the antecedent is not). But this is just the Curry-sentence so it is true as well. Again by modus ponens we arrive at the absurd conclusion.

So in order to block the trivialisation of our theory of truth, we should focus on derivations of triviality rather than mere inconsistencies. As such, we will group them together by a more formal criterion. Both cases can be subsumed under the derivability of the empty sequent $\Rightarrow$. If a paradox leads to a contradiction, we derive $\Rightarrow \phi$ and $\Rightarrow \neg\phi$. If we can derive any formula whatsoever, we can of course also derive such a contradiction. The empty sequent is then easily derived:

$$\dfrac{\begin{array}{c}\vdots\\ \Rightarrow \neg\phi\end{array} \qquad \dfrac{\begin{array}{c}\vdots\\ \Rightarrow \phi\end{array}}{\neg\phi \Rightarrow}}{\Rightarrow}$$

So paradoxes leading to these kinds of unacceptable conclusions can be defined as proofs ending in the empty sequent:

Definition 2.3.1. Paradox

A proof t is paradoxical (or is called a paradox) iff it ends in the empty sequent. We write p to denote a paradox.

Almost all of the paradoxes of truth which are discussed in the literature and which we will discuss here can be subsumed under our formal notion of paradox. But there is one kind of conclusion, which evades this definition but is often seen as unacceptable and thus paradoxical: An ω-inconsistency, i.e. proofs of $\Rightarrow \neg\forall x\phi x$ and $\Rightarrow \phi\overline{n}$ for every $n \in \omega$. The intuitive trouble here is that every instance of ϕx is derivable, yet also the negation of the universal claim that ϕ holds of all instances. From the informal perspective, the universal formula should be true given that all its instances are, which then forms a contradiction with the derivation of $\neg\forall x\phi x$. As such, a contradiction as in the Liar-case is just around the corner.

Formally of course this contradiction cannot be obtained in first-order theories (as our NT) due to Compactness: there is no finite subset of premises which entails the contradiction. This also means that ω-inconsistencies cannot be subsumed under our definition of paradox above. But we will stick to the definition and discuss ω-inconsistencies separately from the other Paradoxes.

2.3.1 All Things Paradoxical

A theory of arithmetic augmented with a truth-predicate T allows us to prove equivalences and identities of what we would informally call paradoxical sentences. We will show below that all truth-theoretic paradoxes typically discussed in the literature can be seen as instances of the weak or strong diagonal lemma:

Theorem 2.3.2. Strong Diagonal Lemma.

Where $\phi x \in \mathcal{L}_T$ having only x free, there is a term t in $\mathcal{L}_T$, s.t. $t = \ulcorner \phi t \urcorner$ is provable in PAT.

Proof. Consider the primitive recursive substitution function $sub(r, s, t)$, which returns a term the code of a formula, which is the result of replacing all free occurrences of s in r by t. We can then define the p.r. diagonalisation function $d(\ulcorner \phi x \urcorner) = sub(\ulcorner \phi x \urcorner, \ulcorner x \urcorner, \ulcorner \phi x \urcorner)$. So given the code of a formula ϕ with only x free, d returns the code of the result of replacing all free occurrences of x by $\ulcorner \phi x \urcorner$. It is then straightforward to show $d(\ulcorner \phi d(x) \urcorner) = \ulcorner \phi(d(\ulcorner \phi x \urcorner)) \urcorner$ by substitution. Letting $d(\ulcorner \phi x \urcorner) = t$ gives us the desired $t = \ulcorner \phi t \urcorner$. □

Theorem 2.3.3. Weak Diagonal Lemma.

Given a formula ϕx with just x free, there is a sentence $\psi \in \mathcal{L}_T$, s.t. $\psi \leftrightarrow \phi \ulcorner \psi \urcorner$ is provable in PAT.

Proof. One way to prove the weak diagonal lemma is via the strong version. Consider an arbitrary formula ϕx with x free. By classical logic it holds that $\phi t \leftrightarrow \phi t$, where t is a term obtained by the strong diagonal lemma s.t. $t = \ulcorner \phi t \urcorner$. By substitution we get $\phi t \leftrightarrow \phi \ulcorner \phi t \urcorner$. It is also possible to prove the weak diagonal lemma without function symbols for p.r. functions such as substitution or diagonalisation in the language. Such proofs accordingly also make no use of the strong diagonal lemma – see e.g. (Picollo, 2018a; Boolos et al., 2007). □

Consider for example the Liar sentence, which informally is a sentence that says about itself that it is not true. By weakly diagonalising the formula $\neg Tx$, we can show that there is a sentence – call it λ – s.t. $\lambda \leftrightarrow \neg T \ulcorner \lambda \urcorner$ is provable in PAT. Using the Strong Diagonal Lemma, we get a variation of the Liar, namely a term l, s.t. $l = \ulcorner \neg Tl \urcorner$ holds.

This section provides derivations for many prominent examples of paradoxes from the literature. The starting point for all of them is the diagonal lemma given above. It sometimes makes a difference in the use of structural rules whether we consider the Weak Diagonal Lemma or the Strong Diagonal Lemma, which is why we always present both versions.

Liar

Consider first the Liar sentence λ obtained from the Weak Diagonal Lemma, for which we can prove the biconditional $\lambda \leftrightarrow \neg T\ulcorner\lambda\urcorner$. Since the sequents $\Rightarrow \lambda \rightarrow \neg T\ulcorner\lambda\urcorner$ and $\Rightarrow \neg T\ulcorner\lambda\urcorner \rightarrow \lambda$ are derivable with Q as a base theory, we can conclude via the inversion lemmata for $\neg$ and $\rightarrow$ that $\lambda, T\ulcorner\lambda\urcorner \Rightarrow$ and $\Rightarrow \lambda, T\ulcorner\lambda\urcorner$ are derivable. The empty sequent then follows swiftly:

$$\dfrac{\dfrac{\begin{matrix}\vdots\\ \Rightarrow \lambda, T\ulcorner\lambda\urcorner\end{matrix}}{\Rightarrow T\ulcorner\lambda\urcorner}\,\mathrm{T2} \qquad \dfrac{\begin{matrix}\vdots\\ \lambda, T\ulcorner\lambda\urcorner \Rightarrow\end{matrix}}{T\ulcorner\lambda\urcorner \Rightarrow}\,\mathrm{T1}}{\Rightarrow}$$

Using the Strong Diagonal Lemma, the Liar is a sentence denoted by a term l, s.t. $l = \ulcorner\neg Tl\urcorner$ holds:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{Tl \Rightarrow Tl}{\Rightarrow Tl, \neg Tl}\,\neg\mathrm{R}}{\Rightarrow Tl, T\ulcorner\neg Tl\urcorner}\,\mathrm{T2}}{l = \ulcorner\neg Tl\urcorner \Rightarrow Tl}\,\mathrm{SubR}}{\Rightarrow Tl}\,{=}E \qquad \dfrac{\dfrac{\dfrac{\dfrac{Tl \Rightarrow Tl}{\neg Tl, Tl \Rightarrow}\,\neg\mathrm{L}}{T\ulcorner\neg Tl\urcorner, Tl \Rightarrow}\,\mathrm{T1}}{l = \ulcorner\neg Tl\urcorner, Tl \Rightarrow}\,\mathrm{SubL}}{Tl \Rightarrow}\,{=}E}{\Rightarrow}$$

An application of Cut to $\Rightarrow Tl$ and $Tl \Rightarrow$ then yields the empty sequent and so a paradox.

Liar Cycle

A Liar Cycle is an ordered, finite list of sentences, s.t. each sentence says about its successor that it is true and the last element says about the first that it is false. As an example we here investigate a Liar Cycle of length two with sentences λ_1, λ_2, where $\lambda_1 \leftrightarrow T\ulcorner\lambda_2\urcorner$ and $\lambda_2 \leftrightarrow \neg T\ulcorner\lambda_1\urcorner$ hold. Using the inversion lemma similar to the Liar case, we can construct the following paradox:

$$\dfrac{\dfrac{\dfrac{\dfrac{\begin{matrix}\vdots\\ \Rightarrow \lambda_2, T\ulcorner\lambda_1\urcorner\end{matrix}}{\Rightarrow T\ulcorner\lambda_2\urcorner, T\ulcorner\lambda_1\urcorner}\,\mathrm{T2} \qquad \begin{matrix}\vdots\\ T\ulcorner\lambda_2\urcorner \Rightarrow \lambda_1\end{matrix}}{\Rightarrow \lambda_1, T\ulcorner\lambda_1\urcorner}\,\mathrm{Cut}}{\Rightarrow T\ulcorner\lambda_1\urcorner}\,\mathrm{T2} \qquad \dfrac{\dfrac{\begin{matrix}\vdots\\ \lambda_1 \Rightarrow T\ulcorner\lambda_2\urcorner\end{matrix} \qquad \dfrac{\begin{matrix}\vdots\\ T\ulcorner\lambda_1\urcorner, \lambda_2 \Rightarrow\end{matrix}}{T\ulcorner\lambda_1\urcorner, T\ulcorner\lambda_2\urcorner \Rightarrow}\,\mathrm{T1}}{T\ulcorner\lambda_1\urcorner, \lambda_1 \Rightarrow}\,\mathrm{Cut}}{T\ulcorner\lambda_1\urcorner \Rightarrow}\,\mathrm{T1}}{\Rightarrow}\,\mathrm{Cut}$$

A strong version of the paradox is achieved by constructing terms $l_1 = \ulcorner Tl_2 \urcorner$ and $l_2 = \ulcorner \neg Tl_1 \urcorner$ via the Strong Diagonal Lemma:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{Tl_1 \Rightarrow Tl_1}{\Rightarrow Tl_1, \neg Tl_1}{\scriptstyle \neg R}}{\Rightarrow Tl_1, T\ulcorner \neg Tl_1 \urcorner}{\scriptstyle T2}}{l_2 = \ulcorner \neg Tl_1 \urcorner \Rightarrow Tl_1, Tl_2}{\scriptstyle SubR}}{\Rightarrow Tl_1, Tl_2}{\scriptstyle =E}}{\Rightarrow Tl_1, T\ulcorner Tl_2 \urcorner}{\scriptstyle T2}}{l_1 = \ulcorner Tl_2 \urcorner \Rightarrow Tl_1}{\scriptstyle SubR}}{\Rightarrow Tl_1}{\scriptstyle =E} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{Tl_1 \Rightarrow Tl_1}{\neg Tl_1, Tl_1 \Rightarrow}{\scriptstyle \neg L}}{T\ulcorner \neg Tl_1 \urcorner, Tl_1 \Rightarrow}{\scriptstyle T1}}{l_2 = \ulcorner \neg Tl_1 \urcorner, Tl_2, Tl_1 \Rightarrow}{\scriptstyle T1}}{Tl_2, Tl_1 \Rightarrow}{\scriptstyle =E}}{T\ulcorner Tl_2 \urcorner, Tl_1 \Rightarrow}{\scriptstyle T1}}{l_1 = \ulcorner Tl_2 \urcorner, Tl_1 \Rightarrow}{\scriptstyle SubL}}{Tl_1 \Rightarrow}{\scriptstyle =E}}{\Rightarrow}$$

Curry

For a weak Curry-sentence κ, we can prove the biconditional $\kappa \leftrightarrow (T\ulcorner \kappa \urcorner \to \phi)$ for some arbitrary $\phi \in \mathcal{L}_T$. Again using the inversion trick gives us the following paradox:

$$\dfrac{\dfrac{\begin{matrix}\vdots \\ \Rightarrow T\ulcorner \kappa \urcorner, \kappa\end{matrix}}{\Rightarrow T\ulcorner \kappa \urcorner}{\scriptstyle T2} \qquad \dfrac{\begin{matrix}\vdots \\ \kappa, T\ulcorner \kappa \urcorner \Rightarrow \phi\end{matrix}}{T\ulcorner \kappa \urcorner \Rightarrow \phi}{\scriptstyle T1}}{\Rightarrow \phi}{\scriptstyle Cut}$$

For the strong version, construct a term k, s.t. $k = \ulcorner Tk \to \phi \urcorner$ holds.

$$\dfrac{\dfrac{\dfrac{\dfrac{Tk \Rightarrow Tk \qquad \phi \Rightarrow \phi}{Tk, Tk \to \phi \Rightarrow \phi}{\scriptstyle \to L}}{Tk, T\ulcorner Tk \to \phi \urcorner \Rightarrow \phi}{\scriptstyle T1}}{k = \ulcorner Tk \to \phi \urcorner, Tk \Rightarrow \phi}{\scriptstyle SubL}}{Tk \Rightarrow \phi}{\scriptstyle =E}$$

$$\vdots$$

$$\dfrac{\dfrac{\dfrac{\dfrac{Tk \Rightarrow \phi}{\Rightarrow Tk \to \phi}{\scriptstyle \to R}}{\Rightarrow T\ulcorner Tk \to \phi \urcorner}{\scriptstyle T2}}{k = \ulcorner Tk \to \phi \urcorner \Rightarrow Tk}{\scriptstyle SubR}}{\Rightarrow Tk}{\scriptstyle =E}$$

Putting the two derivations together via Cut gives us a proof of $\Rightarrow \phi$.

McGee

One much discussed example of a sentence, which generates an ω-inconsistency for (some) theories of truth is the McGee sentence (see (McGee, 1985, 1990)). It may be seen as a universally quantified version of the Liar as it says about itself that not all iterations of the truth-predicate on it are true. To represent this formally, consider the function $f(n, \#\phi)$, which is defined as follows:

$$f(0, \#\phi) = \#\phi$$
$$f(n+1, \#\phi) = \tau f(n, \#\phi)$$

τ is a primitive recursive function which applied to $\#\phi$ yields $\#T\#\phi$. Let $\underset{\cdot}{f}$ be the function symbol in $\mathcal{L}_T$ which represents f. By weak diagonalisation we can then obtain a sentence μ, s.t. $\mu \leftrightarrow \neg\forall x T\underset{\cdot}{f}(x, \ulcorner\mu\urcorner)$ holds. Again, using the inversion trick we get sequents $\Rightarrow \forall x T\underset{\cdot}{f}(x, \ulcorner\mu\urcorner), \mu$ and $\mu, \forall x T\underset{\cdot}{f}(x, \ulcorner\mu\urcorner) \Rightarrow$. Some elaborate reasoning then leads to an ω-inconsistency as follows. First, we derive the case of $\phi\overline{0}$, i.e. $T\underset{\cdot}{f}(\overline{0}, \ulcorner\mu\urcorner)$:

$$\dfrac{\dfrac{\dfrac{\dfrac{\begin{array}{c}\vdots\\ \Rightarrow \forall x T\underset{\cdot}{f}(x, \ulcorner\mu\urcorner), \mu\end{array} \qquad \dfrac{T\underset{\cdot}{f}(0, \ulcorner\mu\urcorner) \Rightarrow T\underset{\cdot}{f}(0, \ulcorner\mu\urcorner)}{\forall x T\underset{\cdot}{f}(x, \ulcorner\mu\urcorner) \Rightarrow T\underset{\cdot}{f}(0, \ulcorner\mu\urcorner)}\,\forall\text{L}}{\Rightarrow T\underset{\cdot}{f}(0, \ulcorner\mu\urcorner), \mu}\,\text{Cut}}{\Rightarrow T\underset{\cdot}{f}(0, \ulcorner\mu\urcorner), T\ulcorner\mu\urcorner}\,\text{T2}}{\ulcorner\mu\urcorner = \underset{\cdot}{f}(0, \ulcorner\mu\urcorner) \Rightarrow T\underset{\cdot}{f}(0, \ulcorner\mu\urcorner)}\,\text{SubR}}{\Rightarrow T\underset{\cdot}{f}(0, \ulcorner\mu\urcorner)}\,=E$$

Starting from an instance $\phi\overline{n}$, we can derive the successor case $\phi S\overline{n}$:

$$\dfrac{\dfrac{\dfrac{\begin{array}{c}\vdots\\ \Rightarrow T\underset{\cdot}{f}(\overline{n}, \ulcorner\mu\urcorner)\end{array}}{\Rightarrow T\ulcorner T\underset{\cdot}{f}(\overline{n}, \ulcorner\mu\urcorner)\urcorner}\,\text{T2}}{\ulcorner T\underset{\cdot}{f}(\overline{n}, \ulcorner\mu\urcorner)\urcorner = \underset{\cdot}{f}(S\overline{n}, \ulcorner\mu\urcorner) \Rightarrow T\underset{\cdot}{f}(S\overline{n}, \ulcorner\mu\urcorner)}\,\text{SubR}}{\Rightarrow T\underset{\cdot}{f}(S\overline{n}, \ulcorner\mu\urcorner)}\,=E$$

Together with the zero-case above, it is clear that $\phi\overline{n}$ for every $n \in \omega$ is derivable. To complete the ω-inconsistency, we show that $\neg\forall x\phi x$ is derivable:

$$
\begin{array}{c}
\vdots \\
\mu, \forall x T\underset{\cdot}{f}(x, \ulcorner \mu \urcorner) \Rightarrow \\ \hline
T\ulcorner \mu \urcorner, \forall x T\underset{\cdot}{f}(x, \ulcorner \mu \urcorner) \Rightarrow \quad \text{T1} \\ \hline
\ulcorner \mu \urcorner = \underset{\cdot}{f}(0, \ulcorner \mu \urcorner), T\underset{\cdot}{f}(0, \ulcorner \mu \urcorner), \forall x T\underset{\cdot}{f}(x, \ulcorner \mu \urcorner) \Rightarrow \quad \text{SubL} \\ \hline
T\underset{\cdot}{f}(0, \ulcorner \mu \urcorner), \forall x T\underset{\cdot}{f}(x, \ulcorner \mu \urcorner) \Rightarrow \quad =E \\ \hline
\forall x T\underset{\cdot}{f}(x, \ulcorner \mu \urcorner) \Rightarrow \quad \forall\text{L} \\ \hline
\Rightarrow \neg \forall x T\underset{\cdot}{f}(x, \ulcorner \mu \urcorner) \quad \neg\text{R}
\end{array}
$$

The same steps can be derived using the Strong Diagonal Lemma for a term $m = \ulcorner \neg \forall x T\underset{\cdot}{f}(x, m) \urcorner$:

$$
\begin{array}{c}
T\underset{\cdot}{f}(0, m) \Rightarrow T\underset{\cdot}{f}(0, m) \\ \hline
\forall x T\underset{\cdot}{f}(x, m) \Rightarrow T\underset{\cdot}{f}(0, m) \quad \forall\text{L} \\ \hline
\Rightarrow T\underset{\cdot}{f}(0, m), \neg \forall x T\underset{\cdot}{f}(x, m) \quad \neg\text{R} \\ \hline
\Rightarrow T\underset{\cdot}{f}(0, m), T\ulcorner \neg \forall x T\underset{\cdot}{f}(x, m) \urcorner \quad \text{T2} \\ \hline
m = \ulcorner \neg \forall x T\underset{\cdot}{f}(x, m) \urcorner \Rightarrow T\underset{\cdot}{f}(0, m), Tm \quad \text{SubR} \\ \hline
\Rightarrow T\underset{\cdot}{f}(0, m), Tm \quad =E
\end{array}
$$

$$
\begin{array}{c}
\vdots \\
\Rightarrow T\underset{\cdot}{f}(0, m), Tm \\ \hline
m = \ulcorner T\underset{\cdot}{f}(0, m) \urcorner \Rightarrow T\underset{\cdot}{f}(0, m) \quad \text{SubR} \\ \hline
\Rightarrow T\underset{\cdot}{f}(0, m) \quad =E
\end{array}
$$

$$
\begin{array}{c}
\vdots \\
\Rightarrow T\underset{\cdot}{f}(\overline{n}, m) \\ \hline
\Rightarrow T\ulcorner T\underset{\cdot}{f}(\overline{n}, m) \urcorner \quad \text{T2} \\ \hline
\ulcorner T\underset{\cdot}{f}(t, m) \urcorner = \underset{\cdot}{f}(S\overline{n}, m) \Rightarrow T\underset{\cdot}{f}(\overline{n}, m) \quad \text{SubR} \\ \hline
\Rightarrow T\underset{\cdot}{f}(S\overline{n}, m) \quad =E
\end{array}
$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\forall x T\dot{f}(x,m) \Rightarrow \forall x T\dot{f}(x,m)}{\neg\forall x T\dot{f}(x,m), \forall x T\dot{f}(x,m) \Rightarrow}\neg\text{L}}{T\ulcorner\neg\forall x T\dot{f}(x,m)\urcorner, \forall x T\dot{f}(x,m) \Rightarrow}\text{T1}}{m = \ulcorner\neg\forall x T\dot{f}(x,m)\urcorner, Tm, \forall x T\dot{f}(x,m) \Rightarrow}\text{SubL}}{Tm, \forall x T\dot{f}(x,m) \Rightarrow}= E}{m = \dot{f}(0,m), T\dot{f}(0,m), \forall x T\dot{f}(x,m) \Rightarrow}\text{SubL}}{T\dot{f}(0,m), \forall x T\dot{f}(x,m) \Rightarrow}= E}{\forall x T\dot{f}(x,m) \Rightarrow}\forall\text{L}}{\Rightarrow \neg\forall x T\dot{f}(x,m)}\neg\text{R}$$

(Fjellstad, 2016b) gives another (much more complex) proof of ω-inconsistency using a McGee-sentence in order to show that the nontransitive approach of (Cobreros et al., 2012) is ω-inconsistent. However, it is much easier to prove the ω-inconsistency without Cut starting from an inconsistency such as $T\ulcorner\lambda\urcorner \wedge \neg T\ulcorner\lambda\urcorner$ as it is provable in our nontransitive approach. Consider the formula $\phi : T\ulcorner\lambda\urcorner \wedge y = y$. We can then prove $\neg\forall x\phi x$ and $\phi\overline{n}$ for all $n \in \omega$ as follows without Cut:

$$\dfrac{\begin{matrix}\vdots \\ \Rightarrow T\ulcorner\lambda\urcorner\end{matrix} \qquad \dfrac{}{\Rightarrow \overline{n} = \overline{n}}=\text{T}}{\Rightarrow T\ulcorner\lambda\urcorner \wedge \overline{n} = \overline{n}}\wedge\text{R} \qquad \dfrac{\dfrac{\dfrac{\begin{matrix}\vdots \\ T\ulcorner\lambda\urcorner \Rightarrow\end{matrix}}{T\ulcorner\lambda\urcorner \wedge \overline{n} = \overline{n} \Rightarrow}\wedge\text{L}}{\forall x(T\ulcorner\lambda\urcorner \wedge x = x) \Rightarrow}\forall\text{L}}{\Rightarrow \neg\forall x(T\ulcorner\lambda\urcorner \wedge x = x)}\neg\text{R}$$

Chapter 3

Reasons Against Transitivity

This chapter gives an overview over different ways of constructing a nontransitive logic and reasons to do so. We begin by exploring some motivations stemming from considerations regarding relevance and go on to reasons regarding paradoxes. Further, we discuss reasons to give up Cut, which come from bilateralism, an inferentialist theory of meaning. Last, we argue for the necessity of restricting Cut rather than giving it up completely in the context of theories of truth over PA and argue that various restrictions from the literature are unsatisfactory.

3.1 Reasons From Relevance

Logicians have been formulating nontransitive logics for different reasons. One line of motivation stems from considerations regarding relevance. The main idea is that logical consequence must make sure that if a conclusion ϕ follows logically from some set of premises Γ, then all premises $\gamma \in \Gamma$ are relevant for obtaining ϕ. What relevance amounts to differs of course between philosophical accounts. However, it is easy to see why transitivity of the logical consequence relation fails in some cases if we demand relevance. Suppose that our relevance constraint demands that for ψ to follow logically from ϕ it must be the case that there is an overlap of propositional variables between ϕ and ψ. Let $\phi \vdash \psi$ and $\psi \vdash \epsilon$ hold. Then it may very well be that there is an overlap of propositional variables between ϕ and ψ as well as between ψ and ϵ but not between ϕ and ϵ. Thus transitivity fails.

One of the earliest (if not the earliest) formulation of a nontransitive logic for reasons of relevance can be found in Bolzano's *Wissenschaftslehre* from 1837. Much later, there is a range of papers exploring alternative failures of transitivity, often motivated by the paradoxes of material implication. These paradoxes, unlike the paradoxes of truth, do not lead to triviality. They rather raise issues with an intuitive or natural language reading of the material conditional $\rightarrow$. For example, it is rather unintuitive that because $1 + 1 = 2$ is true, it follows that 'If pigs can fly, then $1 + 1 = 2$' holds. Again, the intuitive request is that the antecedent of the conditional should somehow be relevant for its consequens. Nontransitive approaches roughly in this line can be found in (Lewy et al., 1958; Smiley, 1958; Epstein, 1979).

These failures of transitivity due to considerations regarding relevance are not important to our project of restricting Cut in formal theories of truth. They are irrelevant for at least two reasons. First, in these approaches, the property of transitivity of the consequence relation is not expressed in terms of a rule in the calculus. Rather, the authors give a new definition of logical consequence, which then turns out to be nontransitive. At least in some cases (see e.g. (Smiley, 1958)), it turns out that the definition of logical consequence works for propositional but not for predicate logic. Second, their relevance conditions would not block an application of Cut to conclude the empty sequent from $\Rightarrow T\ulcorner\lambda\urcorner$ and $T\ulcorner\lambda\urcorner \Rightarrow$. Thus the restrictions do not offer a way around the paradoxes and so no non-trivial subtheory of NT.

However, there are also authors, which develop nontransitive logics on the basis of considerations about relevance specifically in order to escape paradoxes of naive notions of truth and set. (Weir, 2005) discusses a nontransitive logic in order to block truth-theoretic paradoxes such as the Liar and (Weir, 1998, 1999) apply the same approach to set-theoretic paradoxes in naive set theory. But Weir's approach as well is not a restriction of transitivity in our sense by putting a restriction on Cut in a sequent calculus. In his system, the lack of some transitivity is a consequence of restrictions of elimination rules for the connectives such as $\neg$ or $\rightarrow$. It is also not the case that all instances of transitivity are lost – all instances of what he calls 'simple transitivity', i.e. if $\phi \vdash \psi$ and $\psi \vdash \epsilon$ then $\phi \vdash \epsilon$, remain valid. As a consequence of the restriction of the rules governing $\to$, the conditional no longer contracts. In Weir's system, it is no longer the case that $\phi \to (\phi \to \psi) \vdash \phi \to \psi$. In NT, however, it is immediate that no operational rule is affected by restricting transitivity, since this restriction amounts to a restriction of Cut, which is a separate rule. Also contraction for the conditional holds in full generality. So although Weir's consequence relation may be nontransitive in some cases, his approach and ours are rather different[1].

3.2 Reasons From Proof-Theory

There is a historical line of investigations into the structure of paradoxes, which is relevant for our purposes. Instead of sequent calculi, the investigation takes place in systems of natural deduction. Translations between sequent calculi and natural deduction are well-understood and it can be shown that an application of Cut corresponds to a derivation in non-normal form in natural deduction (von Plato, 2003). A derivation is in non-normal form iff it introduces a complex formula and this complex formula is also a premise in an application of an elimination rule of the introduced connective. For example, consider the following two natural deduction rules:

$$\dfrac{\begin{array}{c}[\phi]\\ \vdots\\ \psi\end{array}}{\phi \to \psi}\,{\to}\mathrm{I} \qquad\qquad \dfrac{\phi \qquad \phi \to \psi}{\psi}\,{\to}\mathrm{E}$$

Then a derivation containing the following applications would be in non-normal form:

$$\dfrac{\begin{array}{c}\vdots\\ \phi\end{array} \qquad \dfrac{\begin{array}{c}[\phi]\\ \vdots\\ \psi\end{array}}{\phi \to \psi}\,{\to}\mathrm{I}}{\psi}\,{\to}\mathrm{E}$$

[1]This is acknowledged by Weir himself, comparing his approach to Ripley's in (Weir, 2015).

The derivation of the conditional $\phi \to \psi$ is a redundancy in the overall derivation of ψ for the conditional is introduced before it is eliminated again. One can then define transformation strategies, which transform derivations from non-normal form into normal form. In the case of the conditional, such a transformation strategy has the following form

$$\dfrac{\begin{array}{c}\vdots\\ \phi\end{array} \qquad \dfrac{\begin{array}{c}[\phi]\\ \vdots\\ \psi\end{array}}{\phi \to \psi}\,{\to}\mathrm{I}}{\psi}\,{\to}\mathrm{E} \qquad \rhd \qquad \begin{array}{c}\vdots\\ \phi\\ \vdots\\ \psi\end{array}$$

It can be shown that all derivations of classical logic can be turned into normal form (see Prawitz, 1965), just as every derivation of classical logic in sequent calculus can be turned into one without Cut (given a suitable formulation as explained in the technical preliminaries). However, in the case of paradoxes such transformations fail. To our knowledge, it was first shown in (Prawitz, 1965) that derivations of absurdity from Russell's paradox do not normalise. Applying transformation strategies similar to the one above will, after some repetitions, give back the original derivation in non-normal form. Thus the derivation does not normalise. The same behaviour of derivations of absurdity has been shown for truth-theroetic paradoxes such as the Liar, Liar-cycle, Curry by (Tennant, 1982, 2015). (Tennant, 1995) extends this analysis successfully to Yablo's paradox (see (Yablo, 1993) for the original paradox).

Neither Prawitz nor Tennant argue explicitly that their analysis of paradoxes in terms of non-normalisability should be turned into a solution to the paradoxes. However, there is a long standing proof-theoretic tradition arguing that only proofs in normal form are 'real proofs' (see e.g. (Prawitz, 1974) for a seminal paper). In fact, Tennant has long been arguing for a nontransitive logic in our sense by giving up the rule of Cut (Tennant, 1987, 2017) for reasons other than the paradoxes. So giving up Cut in order to escape triviality in theories of truth seems to be an independently motivated approach. The motivation is in line with the proof-theoretic tradition just mentioned: Proofs in non-normal form (and thus with Cuts) are seen as impure, as they contain roundabouts. In natural deduction, connectives are introduced only then to be eliminated again. Proper logical consequence ought to be defined only by those proofs which are free of such redundancies. So the reason to give up Cut can be summarised as follows:

> **Redundancy Criterion**
>
> Applications of Cut should be avoided completely in order not to allow redundancies in proofs.

3.3 Bilateralist Reasons

David Ripley is currently the most prominent proponent of a nontransitive logic and applies it to a wide range of paradoxes (truth-theoretic and others). A strength of Ripley's approach is that the argument to give up Cut does not stem from the paradoxes directly but has independent grounds in a theory of meaning called bilateralism, which was originally introduced in (Rumfitt, 2000). Bilateralism is an inferentialist theory of meaning. Roughly, this amounts to saying that the meaning of e.g. a logical constant is defined by the rules governing it. This particular inferentialist theory takes two speech acts as primitive: assertion and denial.

A sequent $\Gamma \Rightarrow \Delta$ is interpreted as a position $\{\Gamma, \Delta\}$, in which all formulae $\gamma \in \Gamma$ are asserted and all $\delta \in \Delta$ are denied. Positions can then either be coherent or incoherent. The intuitive idea here is that the consequence claim of the sequent $\Gamma \Rightarrow \Delta$ is reduced to the level of speech acts: $\bigvee \Delta$ follows from $\bigwedge \Gamma$ iff it is incoherent to assert all formulae in Γ but to deny all in Δ.

Sequent rules can then be justified via this reading by demanding that rules lead only from incoherent positions to other incoherent positions as argued in (Restall, 2005, 2009, 2013). For example, consider the rule ¬R:

$$\frac{\phi, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg\phi}\ \neg\text{R}$$

This inference is then justified as follows: Assume that it is incoherent to assert ϕ and $\bigwedge \Gamma$ but to deny all formulae in Δ. If (given the context) it is incoherent to assert ϕ, then it must be incoherent as well to deny its negation, which amounts to introducing $\neg\phi$ on the right-hand side of the sequent. The crucial question now is how the bilateralist story fares with respect to the rule of Cut:

$$\frac{\Gamma \Rightarrow \Delta, \phi \qquad \phi, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}\text{Cut}$$

To see the problem with Cut, consider its contraposition. Then Cut amounts to the claim that if the position $\langle \Gamma, \Delta \rangle$ is coherent, then at least one of the positions $\langle \{\phi\} \cup \Gamma, \Delta\}$ or $\langle \Gamma, \Delta \cup \{\phi\}\rangle$ is coherent as well – for any ϕ. In terms of assertion and denial this amounts to the claim that any position can be extended by either an assertion or a denial of any sentence while remaining coherent. This is why both Restall and Ripley speak of Cut as an *extensibility criterion* of the consequence relation understood in the bilateralist framework.

Such an extensibility criterion seems to be perfectly fine for many domains and in fact it is inevitable for systems in which Cut is an eliminable or an admissible rule: There Cut holds, even if we decide to drop it. So the point is not to argue against Cut and expect it to have counterexamples in every domain. Again, some systems are closed under Cut even if the rule is absent. Rather, Ripley simply points out that the closure under Cut (and thus transitivity) is something that is to be discovered rather than to be imposed on a set of rules:

> [...] we might think that extensibility only *happens* to hold over the usual classical rules. If this is so, there's no particular reason to expect it to continue to hold when we consider a richer set of rules, such as one including appropriate constraints on T. As we have seen, it does not: there are things (like the empty sequent) derivable in [NT] plus cut that are not derivable in [NT] alone. (Ripley, 2013a: p.151f)

The crucial question is when we should expect Cut, the extensibility criterion, to hold and when not. Ripley (as in the passage just cited) seems to suggest that extensibility should be assumed if and only if Cut is an admissible rule:

> **Admissibility Criterion**
>
> Cut should only be applicable where it is admissible.

3.4 Restricting Cut

3.4.1 The Need For Cut

In this section we provide arguments for the claim that despite the issues with Cut, it is necessary to keep some instances of it. We also argue that the Redundancy and Admissibility criterion are not apt to motivate a useful restriction. It will be shown that there are theories and other formulations of classical logic which depend on the presence of Cut in order to be deductively as strong as one would like them to be.

We already saw an example of this in the technical preliminaries for the case of arithmetical theories. Where it is still possible to have Cut as an admissible rule for the rather weak arithmetical theory Q, this is no longer possible if we add Induction in order to obtain full PA. To be more precise, Cut-elimination for PA is not possible in a setting in which we require proofs to be finite objects (i.e. have a finite number of nodes). (Siders, 2013) gives an improved reconstruction of Gentzen's original proof of the consistency of PA via Cut-elimination techniques and (Siders, 2015) shows normalisability of PA in natural deduction (which translates to Cut-elimination in sequent calculus). As a historical note: Gentzen's original proof (or rather all four versions of it) are not proofs of full Cut-elimination for PA but rather transformations of proofs s.t. all Cuts have atomic arithmetical formulae as their Cut-formula. This is enough to conclude consistency.

Despite the normalisability proof of (Siders, 2015), it is crucial to see that some proofs in normal form are of transfinite height. They can only be given in a finite setting if Cut (or non-normal derivations) is (are) allowed. Recall that Ripley argued that Cut as the extensibility criterion on assertion and denial is not to be imposed on the consequence relation but has to be discovered by showing its admissibility. This then fits nicely with paradoxes of truth since Cut is no longer admissible if we extend the system with transparent T-rules. However, we take it that there are no paradoxes in PA. Typically, we think of PA to behave perfectly classical and so the extensibility criterion should hold here as well: for any arithmetical ϕ and every coherent position, that position can be extended by either the assertion or denial of

ϕ s.t. the result is coherent as well. Thus Cut should be admitted as a rule despite the fact that the Redundancy and Admissibility Criterion tell us otherwise. So the admissibility criterion does not match the reading of Cut as an extensibility criterion.

A similar point can be made with respect to even stronger theories in set theory s.a. ZF. Here the situation is even worse. Not only are there no finite proofs for every theorem of ZF without Cut, we even know that Cut-elimination in it fails (despite a reasonable formulation of sequent rules). Crabbé provided a rather straightforward counterexample to the normalisability of ZF in (Crabbé, n.d.). Here the same argument works: In our conversational norms we would like to be able to either assert or deny any ϕ of the language of ZF. However, Cut is not admissible here, so by the Redundancy and Admissibility Criterion, we must not add Cut to our set theory.

Last, we want to give another example of theories in which Cut-admissibility fails, yet we expect that the extensibility criterion to hold. For this purpose, consider again CL, i.e. just the operational rules for $\neg, \rightarrow, \forall$ and the structural rules. With the given formulation of the operational rules, Cut can be shown to be admissible. However, this fact is highly dependent on the formulation of these rules. Consider for example the alternative formulation CL^* in which we replace ¬L by the initial sequent $\phi, \neg\phi \Rightarrow$. Given Cut, we can then easily show ¬L to be admissible:

$$\frac{\Gamma \Rightarrow \Delta, \phi \qquad \phi, \neg\phi \Rightarrow}{\neg\phi, \Gamma \Rightarrow \Delta}\text{Cut}$$

So CL and CL^* with Cut are deductively equivalent. But without Cut, this admissibility fails. So in CL^*, Cut is not eliminable. The Admissibility Criterion would then tell us that we cannot expect extensibility to hold here. However, we are in pure classical first-order logic. There is nothing to be suspicious about when it comes to the fact that every ϕ is either assertable or deniable. So given the examples of PA, ZF and alternative formulations of CL, we conclude that we need a restriction on Cut more sophisticated than checking its admissibility. The following section discusses some proposals of restricted Cut rules from the literature, which we argue have problems as well. The following chapter then introduces new reasons to be suspicious of Cut and how to restrict it accordingly.

3.4.2 Some Attempts To Restrict Cut

One restriction of Cut in the literature is discussed in (Schroeder-Heister, 2003, 2012a,b, 2016) and (Tranchini, 2016) for a natural-deduction formulation of the same technical idea. Schroeder-Heister's notation is slightly different as he is working in a typed sequent calculus, but we can easily adapt the idea to our untyped setting. The motivation to restrict Cut comes from the Prawitz-Tennant tradition discussed above, in which only proofs in normal form or without Cuts are considered to be proper proofs. Schroeder-Heister proposes to restrict Cut, s.t. the rule can only be applied iff the application is eliminable.

There are at least two problems with this proposal. First, we have already argued that there are applications of Cut in theories such as PA, which cannot be eliminated. So to propose that Cut is only applicable where it is eliminable again means to cripple basic arithmetic. Second, the criterion is too complex: Depending on the theory we are working in, judging whether or not an application of Cut is eliminable is not a decidable matter. This definitely holds in the case of arithmetical theories such as PA. So the proposed restriction is too complex in order to work as a restriction on a formal theory such as NT.

Another proposal for a restriction on Cut is given in (Barrio et al., 2018). Here Cut is restricted via a recovery operator $\circ$, inspired by recent developments of LFIs (Logics of Formal Inconsistency) (see Carnielli et al., 2007):

$$\frac{\Gamma \Rightarrow \Delta, \phi \qquad \phi, \Gamma \Rightarrow \Delta \qquad \Rightarrow \circ\phi}{\Gamma \Rightarrow \Delta}\text{ Gentle Cut}$$

The intended reading of the recovery operator is then that $\circ$ applies to ϕ, whenever ϕ can be assumed to receive a classical truth-value (see (Barrio et al., 2018: p.6)). So far so good, but the question remains: how are we to determine which formulae ϕ can safely be assigned a classical truth-value? In an attempt to answer this question, the authors point to Kripke's fixed point construction famously introduced in (Kripke, 1975). There Kripke develops a strategy of how to build three-valued models (given some three-valued valuation scheme) by building transfinite stages on top of a ground model, in which every successor stage puts into the extension of the truth-predicate what is true in the previous model (and analogously for the anti-extension of T). What ends up in the fixed point is called grounded and is thus often regarded to be non-pathological.

However, the authors themselves point out that there are at least two problems when trying to restrict Cut to grounded ϕ. First, what ends up being grounded depends on the choice of the base model. It is hard to see what criteria could determine the choice of the correct base model. Second (if we consider all possible fixed points and not just the set of grounded expressions), Kripke's construction process leads to an infinite plethora of fixed points. Thus one would have to give a criterion to pick out exactly one of these fixed points and the authors admit that they have no such criterion available. One may further point out that what ends up being grounded is also dependent on the valuation scheme. Kripke himself shows that the construction can be executed using the strong Kleene scheme, weak Kleene scheme or a supervaluationist scheme. So one would have to give additional arguments about which valuation scheme to choose in order to determine the grounded sentences.

As a consequence of such problems, Barrio et al. conclude that "[...] there is no absolutely general way of determining which are the non-pathological sentences according to Kleene-Kripke valuations and, thus, of determining which sentences ought to be marked with the recovery operator." (Barrio et al., 2018: p.7) But even given that the above questions are satisfactorily answered, there remains a bigger issue about the proposal which cannot be overcome. Kripke's fixed points are arithmetically very complex: the minimal fixed point is known to be too complex to be even recursive and so restricting Cut to grounded ϕ is not decidable. The arithmetical part of the fixed point model is of course not axiomatisable by Gödel's first incompleteness theorem. (Fischer et al., 2015) show that the truth-theoretic part of the model is not axiomatisable either. Thus the authors are confronted with the burden to show an axiomatic way of approaching the truth-theoretic part of the fixed point model.

The authors go on to explain that these issues do not need to end the project of recovering applications of Cut in terms of $\circ$. They argue that "[...] although we cannot exhaustively pin down a set of sentences that can be marked with the recovery operator, we can still recover all the meta-inferences [...] by *reasoning* *under the assumption* that some of the featured sentences are non-pathological“ (Barrio et al., 2018: p.7). However, they say little to nothing about the notion of assumption at hand here. The most pressing questions are left unanswered: Can we assume any ϕ to be non-pathological? What happens if our assumption turns out to be false as in the case of the Liar?[2] We conclude that this approach would have to clarify the relevant notion of assumption here before it can be properly assessed.

If the above problems weren't enough, Barrio et al. deliver the striking blow against the applicability of their approach to our setup of self-reference in NT. (Barrio et al., 2018: p.10) shows that if self-reference in the underlying theory is expressed in terms of identity, then the nontransitive system with recovered Cut via $\circ$ is trivial. NT clearly fulfils this criteria, for it is strong enough to express self-reference in terms of identity via the strong diagonal lemma and thus proves the existence of Liars like $l = \ulcorner \neg Tl \urcorner$. Thus we cannot recover instances of Cut in NT via $\circ$ – at least not in the way proposed by Barrio et al. Our formal theory of truth must be able to handle paradoxical sentences expressed as identities.

At this point, one might wonder whether we simply do not choose a very easy, pragmatic and straightforward way of restricting Cut: It is clear that we cannot have Cut for the full language $\mathcal{L}_T$, but we can have it for the full arithmetical language $\mathcal{L}_{\mathsf{PA}}$. This would fit neither the Redundancy, nor the Admissibility Criterion, since Cut is not eliminable in PA. However, one might argue that this is a suitable restriction because by the (at least plausible) consistency of PA, we can safely apply Cut whenever the Cut-formula is contained in $\mathcal{L}_{PA}$. Let $\mathsf{NT}_{\mathsf{PA}}$ be the result of adding to NT the rule of Cut given that the Cut-formula $\phi \in \mathcal{L}_{\mathsf{PA}}$.

It is easy to see that $\mathsf{NT}_{\mathsf{PA}}$ is non-trivial (assuming the consistency of PA). However, we want to argue that it is too weak to be a satisfactory theory of truth. Restricting Cut to only arithmetical formulae has the consequence that we can do very little with the truth predicate. By this we mean that $\mathsf{NT}_{\mathsf{PA}}$ may prove all instances of the (uniform) T-schema, but it does not allow us to prove new arithmetical theorems.

To be fair, some authors have argued that a theory of truth ought to be

[2]The authors cite (Carnielli and Coniglio, 2016: p.18), but the cited source does not help in clarifying the notion of assumption here.

conservative over its base theory, i.e. it precisely should not allow one to prove any new arithmetical theorems when added to PA (see e.g. Cieslinski, 2017). However, the non-conservativity is also often seen as a desideratum (see e.g. Leitgeb, 2007) of theories of truth (we will discuss this in more detail below). Many (mostly classical) theories of truth over PA are strong enough to prove the consistency of PA in terms of a purely arithmetical formula. We can easily show that $\mathsf{NT}_{\mathsf{PA}}$ proves no such thing, i.e. it is $\mathcal{L}_{\mathsf{PA}}$-conservative.

Definition 3.4.1. $\mathcal{L}_{PA}$-conservativity

A theory S is said to be $\mathcal{L}_{PA}$-conservative over a theory S' iff for all $\Gamma, \Delta \subseteq \mathcal{L}_{PA}$ it holds that $\vdash_S \Gamma \Rightarrow \Delta$ iff $\vdash_{S'} \Gamma \Rightarrow \Delta$.

Lemma 3.4.1. $\mathsf{NT}_{\mathsf{PA}}$ is $\mathcal{L}_{\mathsf{PA}}$-conservative over PA.

Let $\Gamma, \Delta \subseteq \mathcal{L}_{\mathsf{PA}}$. If $\vdash_{\mathsf{NT}_{\mathsf{PA}}} \Gamma \Rightarrow \Delta$, then $\vdash_{\mathsf{PA}} \Gamma \Rightarrow \Delta$

Proof. Assume for contradiction that there is an arithmetical sequent $\Gamma \Rightarrow \Delta$ which is provable in $\mathsf{NT}_{\mathsf{PA}}$ but not in PA alone. Since it is not provable in PA, the proof of the sequent must contain some instances of T1 or T2, which introduce a T-formula. However, since the conclusion sequent is purely arithmetical, the principal T-formula must have been eliminated, which is impossible without Cut for formulae containing T. □

So although restricting Cut to arithmetical formulae works in the sense of leading to a nontrivial theory of truth arithmetically as strong as PA, it still leaves Cut and thus the proof-theoretic power rather crippled. Our goal is to find a nontransitive theory of truth (or multiple such theories) which overcome all of the above problems: A nontransitive theory of truth with a strong Cut-rule and which is proof-theoretically strong. To do so, we will formulate a restriction on Cut in a rather unusual way. Instead of finding some recursive subset of $\mathcal{L}_T$, s.t. we can safely apply Cut as long as the Cut-formula is contained in that subset, we will choose a restriction relative to the rules used so far in the proof. The following chapter shows that Cut is a rule, which implicitly assumes the consistency of the subderivations leading to its premises. On the basis of this observation, we restrict Cut to cases in which the consistency of these premises is guaranteed.

Chapter 4

Cut and Consistency

In the previous chapter we have seen some reasons to restrict Cut and how to do so, as well as arguments why these reasons and ways of restricting Cut are unsatisfactory. In this chapter, we present a new motivation to restrict Cut and discuss different ways of translating this motivation into axiomatisable restrictions. This new restriction is based on the observation that Cut includes an implicit assumption of consistency. Thus Cut should only be applicable to sequents derived in a consistent theory.

Generally, when one wants to restrict a rule, one has to answer at least two questions: First, why should the rule be restricted? Second, how should it be restricted? Bilateralism gives us an interesting and intuitive answer to the first question: Cut amounts to the assumption that every coherent position can be coherently extended by either the assertion or the denial of any formula ϕ. But such a principle is questionable in the presence of paradoxical sentences such as the Liar. This answer to the first question translates into the following answer to the second question on how the restriction looks like: Cut can be applied iff the position can be safely extended by either the assertion or denial of any ϕ. But how is this supposed to be axiomatised? Ripley's answer is that we can only apply Cut if the system proves its conclusion anyway: The safe instances of Cut are its admissible ones. We already saw that this answer is unsatisfactory since Cut is not admissible in many theories where we would expect the extensibility criterion to hold, including PA.

First, we show that Cut is equivalent to a rule expressing an assumption of consistency. This is achieved by showing that a rule expressing such an assumption in natural deduction can be translated into a sequent rule and is equivalent to Cut. This analysis of Cut as a rule expressing an assumption of consistency suggests that Cut may only be applied in a consistent context. The rest of the chapter discusses two ways of axiomatising this approach. The first comes from default logics but turns out to be inapplicable. Lastly, we sketch my own approach, which is based on the idea that in order for Cut to be applicable, its premises need to be derived in a theory known to be consistent.

4.1 A Rule Assuming Consistency

The goal here is to show that Cut is equivalent to a rule expressing an assumption of consistency. Typical formulations of first-order logic or PA do not include such a rule. So how should a rule expressing such a consistency assumption look like? We understand consistency as the absence of a contradiction. A contradiction is any formula of the form $\phi \wedge \neg\phi$. Its absence can be expressed via negation, so $\neg(\phi \wedge \neg\phi)$. Classically this is provably equivalent to $\phi \vee \neg\phi$, the law of excluded middle. Thus we can understand consistency on the level of our object language in terms of the law of excluded middle.

How then can we understand the *assumption* of consistency as a sequent rule? Sequent calculi do not include notions of assumption but there are other calculi, such as natural deduction, which do and which have a well understood connection to sequent calculi. In natural deduction, one can make assumptions within derivations. Certain rules then allow one to discharge these assumptions, concluding some formulae, s.t. its derivation is no longer dependent on the original assumption. To recall how this works, consider as an example the following standard natural deduction rule for the introduction of the material conditional:

$$\dfrac{\begin{matrix}[\phi]^1 \\ \vdots \\ \psi\end{matrix}}{\phi \to \psi}\ {\to}\mathrm{I}^1$$

Given a proof of ψ from the assumption of ϕ (and possibly other assumptions and subderivations), the rule allows us to drop the assumption of ϕ. The derivation of the conditional $\phi \to \psi$ is no longer dependent on the assumption of ϕ. Given this schema and an understanding of consistency in terms of the law of excluded middle, we can construct a natural deduction rule expressing the assumption of consistency as follows:

$$\cfrac{\begin{array}{c}[\phi \vee \neg\phi]^1 \\ \vdots \\ \psi\end{array}}{\psi}\ \text{EME}^1$$

Given a derivation of ψ from the assumption of $\phi \vee \neg\phi$, we can infer ψ and drop the assumption of $\phi \vee \neg\phi$. Whatever ψ we have derived assuming consistency, its derivability is no longer dependent on that assumption. The use and discharge of assumptions in natural deduction is found in terms of left-hand rules in sequent calculi. The rule for the conditional above corresponds to the right-hand rule for the conditional in NT:

$$\frac{\phi, \Gamma \Rightarrow \Delta, \psi}{\Gamma \Rightarrow \Delta, \phi \to \psi}\ {\to}\text{R}$$

Given a proof from ϕ to ψ (ignoring the context), we can conclude the conditional $\phi \to \psi$ while the occurrence of ϕ on the left-hand side disappears – just like when it is discharged in the case of the natural deduction rule. So given a natural deduction rule discharging an assumption of consistency and this translation between rules of natural deduction and sequent calculi, our sequent rule expressing an assumption of consistency must look like this:

$$\frac{\phi \vee \neg\phi, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}\ \text{EME}$$

Where EME stands for Excluded Middle Elimination. We now only need to show that EME and Cut are equivalent:

Lemma 4.1.1. Cut and EME are interderivable.

Proof. For the purpose of this proof, consider the following standard left rule for disjunction:

$$\frac{\phi, \Gamma \Rightarrow \Delta \qquad \psi, \Gamma \Rightarrow \Delta}{\phi \lor \psi, \Gamma \Rightarrow \Delta}\ \lor\text{L}$$

This rule of course correspond to the rules for $\neg$ and $\rightarrow$ in NT given the expressibility of disjunction in terms of negation and the conditional.

We first show that given Cut, EME is a derivable rule:

$$\frac{\dfrac{\dfrac{\phi \Rightarrow \phi}{\Rightarrow \phi, \neg\phi}\ \neg\text{R}}{\Rightarrow \phi \lor \neg\phi}\ \lor\text{R} \qquad \begin{array}{c}\vdots\\ \phi \lor \neg\phi, \Gamma \Rightarrow \Delta\end{array}}{\Gamma \Rightarrow \Delta}\ \text{Cut}$$

Second we show that Cut is derivable given EME:

$$\frac{\dfrac{\dfrac{\begin{array}{c}\vdots\\ \Gamma \Rightarrow \Delta, \phi\end{array}}{\neg\phi, \Gamma \Rightarrow \Delta}\ \neg\text{L} \qquad \begin{array}{c}\vdots\\ \phi, \Gamma \Rightarrow \Delta\end{array}}{\phi \lor \neg\phi, \Gamma \Rightarrow \Delta}\ \lor\text{L}}{\Gamma \Rightarrow \Delta}\ \text{EME}$$

Alternatively, we can also prove this fact by observing that the sets of premises of the rules are interderivable. From the Cut premises $\Gamma \Rightarrow \Delta, \phi$ and $\phi, \Gamma \Rightarrow \Delta$ we get $\phi \lor \neg\phi, \Gamma \Rightarrow \Delta$ by first applying $\neg$L to $\Gamma \Rightarrow \Delta, \phi$ to get $\neg\phi, \Gamma \Rightarrow \Delta$ and then $\lor$L with the other Cut premise to introduce the disjunction. The other direction from the EME premise $\phi \lor \neg\phi, \Gamma \Rightarrow \Delta$ to the Cut premises follows by the invertibility of $\neg$L and $\lor$L. For a proof of the invertibility of $\lor$L, see (Negri et al., 2008: p.32). □

Thus Cut is equivalent to EME, a rule expressing an assumption of consistency. So Cut is a rule expressing an assumption of consistency. The same argument can be made with a similar rule, which we call contradiction elimination (CE):

$$\frac{\Gamma \Rightarrow \Delta, \phi \land \neg\phi}{\Gamma \Rightarrow \Delta}\ \text{CE}$$

CE allows one to eliminate a contradiction on the right-hand side of the sequent. The reasoning behind this rule could be described as follows: The right-hand side of a sequent is read disjunctively, so the contradiction on the far right is merely a disjunct (given that Δ is non-empty). Assuming consistency means to assume the absence of a contradiction, which is its negation alas $\neg(\phi \wedge \neg\phi)$. Via a disjunctive syllogism, we can then drop the disjunct which is the contradiction, leaving us with the rest of the disjunction made of the formulae in Δ. But this reasoning of course only works given the assumption of consistency in the form of $\neg(\phi \wedge \neg\phi)$. So CE as well includes an implicit assumption of consistency. It is easy to show that CE is interderivable with EME and thus with Cut. This further supports my argument that Cut as a rule makes an implicit assumption of consistency.

Lemma 4.1.2. CE, EME and Cut are interderivable.

Proof. Given the lemma above, it suffices to show that CE and EME are interderivable. Given a proof of $\phi \vee \neg\phi, \Gamma \Rightarrow \Delta$, it follows by inversion of $\vee$L and $\neg$L that the sequents $\Gamma \Rightarrow \Delta, \phi$ and $\phi, \Gamma \Rightarrow \Delta$ are provable. An application of $\neg$R gives us $\Gamma \Rightarrow \Delta, \neg\phi$ and using the standard right rule for conjunction

$$\frac{\Gamma \Rightarrow \Delta, \phi \qquad \Gamma \Rightarrow \Delta, \psi}{\Gamma \Rightarrow \Delta, \phi \wedge \psi} \wedge\mathrm{R}$$

then concludes $\Gamma \Rightarrow \Delta, \phi \wedge \neg\phi$. The other direction works analogously: Assuming a proof of $\Gamma \Rightarrow \Delta, \phi \wedge \neg\phi$, the invertibility of $\wedge$R ensures that we have proofs of $\Gamma \Rightarrow \Delta, \phi$ and $\Gamma \Rightarrow \Delta, \neg\phi$. $\neg$L and inversion of $\neg$R give us sequents $\phi, \Gamma \Rightarrow \Delta$ and $\neg\phi, \Gamma \Rightarrow \Delta$. An application of $\vee$L finally concludes $\phi \vee \neg\phi, \Gamma \Rightarrow \Delta$. So the premises of CE and EME are interderivable, which implies that the rules themselves are interderivable as well. Again, since Cut and EME are shown to be interderivable, this suffices to conclude that CE and Cut are interderivable as well. □

Given the analysis of Cut as a rule making an implicit assumption of consistency, the reason to restrict Cut is simple: A rule should only be applicable if its (implicit) assumptions are true. In the case of NT, however, we know the assumption of consistency to be false – for we can easily derive a contradiction. This answer to the question of why the rule should be restricted also leads to an answer to the question of what the restriction should look like. Cut should only be applied in consistent contexts. The following subsections explore how this restriction can be made precise in terms of axiomatisable restrictions on Cut.

But before we go on to discuss potential axiomatisations of this so far informal idea of restricting Cut, we want to note a further benefit of this motivation in contrast to the bilateralist argument. The argument against the unrestricted applicability of Cut from Ripley rests on an understanding of logical consequence in terms of assertion, denial and (in-)coherent positions. Thus this motivation to restrict Cut is only appealing to bilateralists or those who are ready to understand logical consequence in bilateralist terms. As such, the scope of the argument against Cut in its general form is rather limited.

This not the case with my argument above. We have made no use of vocabulary specific to any kind of particular philosophical theory or understanding of logical consequence. The only, also independently motivated, assumptions we made were that consistency in the object language can be understood as the negation of a contradiction and that assumptions are represented in sequent calculi by rules in which formulae on the left-hand side disappear. The latter was shown to directly correspond to rules making use of assumptions in natural deduction. So my argument is much wider in scope and generally applicable in discussions about Cut or suitable restrictions to get around paradoxes.

4.2 Default Logics

Probably the most well-understood and most prominent approach of putting a consistency constraint on rules is to make use of a default logic. A default logic is a logical system including default rules, which are of the form

$$\frac{\phi : \epsilon}{\psi}$$

In such a default rule, ϕ is called the prerequisite, ϵ the justification (or default) and ψ the conclusion. One interpretation is that given ϕ and no information that ϵ

is false, one can conclude ψ (see (Antonelli, 2005: p.14)). Another interpretation makes the idea of retaining consistency via default rules more explicit: If ϕ is provable and $\neg\epsilon$ is not provable, then infer ψ (see (Brewka et al., 1997: p.41)). The latter interpretation by talking about (un)provability of formulae sounds already pretty close to what we would expect in a sequent rule, so we will go with this one. Using this idea for a restriction of Cut, we would, ideally, like to restrict Cut as follows: Given proofs of $\Gamma \Rightarrow \Delta, \phi$, $\phi, \Gamma \Rightarrow \Delta$ and the unprovability of $\phi \wedge \neg\phi$, conclude $\Gamma \Rightarrow \Delta$.

How could such default rules be formulated in sequent calculi? (Bonatti and Olivetti, 1997) offer a rather straightforward way of doing so. In order to express the justification (or the unprovability claim) of a default rule, the sequent system is augmented by anti-sequents $\Gamma \not\Rightarrow \Delta$. Where sequents express a claim of consequence between the left- and right-hand side, anti-sequents express the opposite: $\bigvee \Delta$ does not follow from $\bigwedge \Gamma$. In the special case of an empty left-hand side $\not\Rightarrow \phi$ we speak of ϕ as an anti-theorem. So our restricted rule of Cut would look like this

$$\frac{\Gamma \Rightarrow \Delta, \phi \qquad \phi, \Gamma \Rightarrow \Delta \qquad \not\Rightarrow \phi \wedge \neg\phi}{\Gamma \Rightarrow \Delta} \text{ DefCut}$$

An Initial sequent of the anti-sequent calculus is then any sequent $\Gamma \not\Rightarrow \Delta$, where Γ and Δ share no atomic formulae and anti-sequents are proven by rules for the logical constants including the following:

$$\frac{\Gamma \not\Rightarrow \Delta, \phi}{\neg\phi, \Gamma \not\Rightarrow \Delta} \qquad \frac{\phi, \Gamma \not\Rightarrow \Delta}{\Gamma \not\Rightarrow \Delta, \neg\phi}$$

$$\frac{\Gamma \not\Rightarrow \Delta, \phi}{\phi \to \psi, \Gamma \not\Rightarrow \Delta} \qquad \frac{\psi, \Gamma \not\Rightarrow \Delta}{\phi \to \psi, \Gamma \not\Rightarrow \Delta}$$

$$\frac{\phi, \Gamma \not\Rightarrow \Delta, \psi}{\Gamma \not\Rightarrow \Delta, \phi \to \psi}$$

Such an anti-sequent calculus can then be shown to be complete with respect to the anti-theorems of classical, propositional logic (see also (Varzi, 1990, 1992) for discussion of such calculi). Thus they nicely and fully capture the intended meaning of anti-theorems. However, it can easily be seen that this strategy is unsuccessful in formulating an interesting restriction of Cut in NT.

The first reason is that a complete axiomatisation of anti-theorems of a theory can only work in principle if it is decidable. For decidability simply means that there is a mechanical procedure to determine both what is and what is not provable. Classical propositional logic is decidable, so the strategy works nicely. However, already adding quantifiers by moving on to predicate logic makes the decidability impossible – let alone considering arithmetical theories. One might object that the anti-sequent calculus need not be complete with respect to the anti-theorems. However, we could then only construct such a calculus for the decidable core of our theory, i.e. something not much more powerful than predicate logic with restricted quantifiers. Since the Cut-rule would then have to be defined relative to this anti-sequent calculus, our Cut rule would be too weak to be interesting.

The second reason is concerned with the question, relative to which theory our anti-sequent calculus is supposed to prove the anti-theorems. An anti-sequent calculus is supposed to prove what is unprovable in some other calculus. However, NT is trivial, so there are no sequents, which it does not prove. So there is nothing about NT for the anti-sequent calculus to axiomatise. We would have to look at non-trivial subtheories of NT which then have a non-empty set of anti-theorems. But then we would not capture the relevant notion of consistency, which is the one relative to NT. Further, the considerations regarding the low complexity of decidable theories would still push us to pick a very weak subtheory of NT. All in all, expressing the consistency statement via an anti-sequent calculus is doomed due to issues of complexity.

However, we will show that there is another way of expressing the non-triviality requirement on Cut. The non-triviality requirement will be formalised in terms of whether or not the Cut-premises have been derived in a consistent theory. The consistency assumption contained in Cut has a local character: it is only concerned with the consistency of the Cut-formula but not necessarily with the consistency of all formulae contained in the contexts of the Cut premises. The new requirement we just sketched – that the Cut premises must be derived in a consistent theory – has a global character. Its consistency notion is concerned not only with a single formula but with all formulae in the subderivations leading up to the Cut premises. The difference between the local and global character also means that the new consistency requirement is considerably stronger than the local one speaking of assumptions. This also entails that if the sequents are derived in a consistent theory, then they do not contain inconsistencies in the local sense and thus the consistency assumption is true.

Note also the crucial difference that the anti-sequent calculus used a notion of being *derivable in* S, whereas we only require a notion of being *derived in* S.

We would like to point out that David Ripley mentions a similar idea of restricting Cut in his work (although he never, to our knowledge, formulated a nontransitive theory of truth with such a restriction):

> This suggests adding a restricted rule of cut to the target systems. Such a rule could be used in a derivation only *above* truth [...] rules, never below them. That is, the premises of the cut would have to be derived without use of the truth [...] rules, although any rules at all could be freely applied to the conclusion of the cut. (Ripley, 2013b: p.11)

Ripley also seems to think that it's not worth fully articulating this restriction, since "[...] cut is admissible for G1c [a sequent calculus for classical logic, i.e. his base theory], it follows that this restricted cut is admissible for the target systems" (Ripley, 2013b: p.11). However, given that Cut is not eliminable (or admissible) in our base theory PA, such a restricted Cut rule would not be admissible in our theory of truth either.

4.3 $\mathcal{L}$- and S-classicality

This kind of restriction turns out to be very powerful in terms of keeping classical logic around. We will see later on that a nontransitive theory with such a restriction on Cut can emulate classical logic in a number of interesting and important ways. One of these is that it proves all classically provable sequents for the full language $\mathcal{L}_T$. This is not the case for any other non-classical approach since, by definition, the classical rules must be dropped for at least some instances involving the truth predicate. To make more sense of what is going on in such cases, we introduce two notions of what it may mean for a theory to be classical or to preserve classical logic. These are the notions of $\mathcal{L}$-classicality and S-classicality:

Definition 4.3.1. $\mathcal{L}$-classicality

Given a language $\mathcal{L}$ and a theory S, S is said to be $\mathcal{L}$-classical iff: if $\vdash_{CL} \Gamma \Rightarrow \Delta$ then $\vdash_S \Gamma \Rightarrow \Delta$, where $\Gamma, \Delta \subset \mathcal{L}$.

Definition 4.3.2. S-classicality

Given theories S, S', where S is closed under classical logic, S' is called S-classical iff: if $\vdash_S \Gamma \Rightarrow \Delta$ then $\vdash_{S'} \Gamma \Rightarrow \Delta$.

When logicians defend the preservation of classical logic, they typically seem to have the notion of $\mathcal{L}$-classicality in mind, where $\mathcal{L}$ is any language . The driving argument against a revision of classical logic is that logic is universal in the sense of being language independent. In the study of pure logic, we derive theorems and use inferences which are valid regardless of the language one may choose to instantiate these theorems with. Non-classical approaches to paradoxes break this idea because they make such theorems and inferences dependent on the language. According to standard non-classical approaches, such inferences can only be used for the arithmetical language but not for formulae involving the truth predicate. So the resulting non-classical theory is not $\mathcal{L}$-classical (and therefore also not S-classical).

Since the universality and independence from language of classical logic is captured by $\mathcal{L}$-classicality, what could be an argument for S-classicality? As we will see in more detail below, for theories of truth to be closed under classical logic often brings with it the benefit of greater proof-theoretic strength. This strength expresses itself in the fact that the theory is able to prove higher ordinals to be well-ordered and to emulate more second-order quantification (more on this below). Such points have been used in the recent literature to argue against non-classical variants of classical theories of truth (see e.g. Halbach and Nicolai, 2018). We agree with such arguments in favour of classical logic and we will show later on that our nontransitive approach is able to handle such requirements just as well the strongest classical theory in these respects.

However, there is a crucial difference between the two notions of classicality: Where there can be theories which are fully $\mathcal{L}$-classical with respect to the language

$\mathcal{L}_T$, this is not possible in the case of S-classicality. For simply let S be NT. Then the claim for S-classicality amounts to demanding that our theory be trivial, which is absurd. So if the notion of S-classicality is to be fruitful at all, we need to be careful relative to which classical theory S we judge the S-classicality of our theory. Later on we will see that our nontransitive approach is (or can be made) S-classical with respect to any classical theory of truth S, i.e. a consistent theory of truth obtained from NT by restricting the rules T1, T2 or the compositional principles.

Chapter 5

Nontransitive Transparency

In the previous chapter we showed that Cut makes an implicit consistency assumption and argued that Cut should accordingly be restricted to cases where this assumption is true. Default logics proved to be ill-suited for our purposes. However, it seemed promising to impose on Cut that it be applicable if and only if both its premises have been derived in a consistent theory of truth. For this captures the requirement that there are no formulae ϕ and $\neg\phi$ in the Cut premises. This chapter fills in the details, shows how to build non-trivial, nontransitive theories of truth and explores some of their properties. We start with a nontransitive theory NTT, which demands that its premises must be derived in PAT and then later strengthen it by considering extensions of PAT.

5.1 The Theory NTT

We can now formulate theories of truth where Cut is restricted to cases where both premises have been derived in a consistent theory of truth. We begin with a radical version of this restriction, using PAT, our weakest theory of truth. Recall that PAT is simply PA formulated in $\mathcal{L}_T$ and contains instances of neither T1, T2 nor the compositional principles. We will call this theory Non-Transitive Transparency (NTT).

Definition 5.1.1. Non-Transitive Transparency NTT

The theory NTT is obtained from NT by replacing its rule of Cut by

$$\frac{\Gamma \Rightarrow \Delta, \phi \qquad \phi, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}\text{ PAT-Cut}$$

where both premises of PAT-Cut must have been obtained using only (instances of) rules of PAT.

First some basic facts about the non-triviality, inconsistency and ω-inconsistency of NTT :

Lemma 5.1.1. NTT is i) inconsistent and ii) ω-inconsistent.

Proof. i) can be easily shown via e.g. the Liar-Paradox that both $\Rightarrow T\ulcorner\lambda\urcorner$ and $\Rightarrow \neg T\ulcorner\lambda\urcorner$ are provable, where all. ii) The ω-inconsistency of NTT cannot be established by the derivations from the McGee-sentence, since both versions (stemming from either the weak or strong diagonal lemma) contain at least one application of Cut, s.t. its premises cannot be derived in PAT alone. However, there is an easy derivation from the inconsistency of the Liar. Consider the formula $\phi : T\ulcorner\lambda\urcorner \wedge y = y$. We can then prove $\Rightarrow \phi\overline{n}$ for every $n \in \omega$ and $\Rightarrow \neg\forall x\phi x$ in NTT:

$$\dfrac{\begin{matrix}\vdots\\ \Rightarrow T\ulcorner\lambda\urcorner\end{matrix} \qquad \dfrac{}{\Rightarrow \overline{n} = \overline{n}}{=}\mathrm{T}}{\Rightarrow T\ulcorner\lambda\urcorner \wedge \overline{n} = \overline{n}}\wedge\mathrm{R} \qquad\qquad \dfrac{\dfrac{\dfrac{\begin{matrix}\vdots\\ T\ulcorner\lambda\urcorner \Rightarrow\end{matrix}}{T\ulcorner\lambda\urcorner \wedge \overline{n} = \overline{n} \Rightarrow}\wedge\mathrm{L}}{\forall x T\ulcorner\lambda\urcorner \wedge x = x \Rightarrow}\forall\mathrm{L}}{\Rightarrow \neg\forall x T\ulcorner\lambda\urcorner \wedge x = x}\neg\mathrm{R}$$

□

Lemma 5.1.2. NTT does not prove the empty sequent.

Proof. Assume for contradiction that the empty sequent is provable in NTT. By the consistency of PAT and its closure under Cut, it follows that PAT does not prove the empty sequent. So the derivation of the empty sequent must contain instances of T1, T2 or the compositional principles. However, Cut would be the only rule of NTT to derive the empty sequent, which is impossible after the use of truth-rules. □

Theorem 5.1.3. NTT is non-trivial.

Proof. Assume for contradiction that NTT is trivial, so it proves all sequents of the form $\Rightarrow \phi$ and $\phi \Rightarrow$. Consider the derivations of $\Rightarrow 0 = 1$ and $0 = 1 \Rightarrow$. Since PAT (by its consistency) does not prove $\Rightarrow 0 = 1$, this derivation must involve some application of truth rules. However, the conclusion is arithmetical so the principal formula of this T-inference must have been eliminated. But the only rule which could do so is Cut, which could not have been applied since the alleged Cut-premises cannot be derived in PAT alone due to the use of truth rules. □

Despite the radical restriction of blocking all Cuts which would happen after making use of a truth rule, it is easy to show that full PAT is preserved in our nontransitive theory and thus all classically provable sequents for the full language $\mathcal{L}_T$ remain provable:

Lemma 5.1.4. If $\vdash_{\mathsf{PAT}} \Gamma \Rightarrow \Delta$ then $\vdash_{\mathsf{NTT}} \Gamma \Rightarrow \Delta$.

Proof. Immediate, since no proof of PAT includes an application of a truth rule. □

Corollary 5.1.4.1. For any $\Gamma, \Delta \subseteq \mathcal{L}_T$: if $\vdash_{CL} \Gamma \Rightarrow \Delta$, then $\vdash_{NTT} \Gamma \Rightarrow \Delta$ (where CL is pure first-order logic).

Proof. Immediate by the previous lemma. □

Thus all classically provable sequents for the full language $\mathcal{L}_T$ remain provable, including all theorems of classical logic. Further NTT is also strong with respect to its truth-theoretic content. We can show that the theory proves all instances of the T-schema and is transparent by showing first the admissibility of the elimination rules for T:

Lemma 5.1.5. T3 and T4 are admissible in NTT.

$$\frac{T\ulcorner\phi\urcorner, \Gamma \Rightarrow \Delta}{\phi, \Gamma \Rightarrow \Delta}\text{T3} \qquad \frac{\Gamma \Rightarrow \Delta, T\ulcorner\phi\urcorner}{\Gamma \Rightarrow \Delta, \phi}\text{T4}$$

are admissible in NTT.

Proof. The proof is the same as for ST in 2.2.4. □

Corollary 5.1.5.1. NTT is a transparent theory of truth

Proof. Immediate by the unrestricted rules T1 and T2, as well as the admissibility of T3 and T4. □

Lemma 5.1.6. For all $\phi \in \mathcal{L}_T, \vdash_{\mathsf{NTT}} \Rightarrow \phi \leftrightarrow T\ulcorner\phi\urcorner$.

Proof. Immediate by NTT's transparency and the unrestricted $\rightarrow$R-rule:

$$\dfrac{\dfrac{\phi \Rightarrow \phi}{T\ulcorner\phi\urcorner \Rightarrow \phi}\,\text{T1}}{\Rightarrow T\ulcorner\phi\urcorner \rightarrow \phi}\,\rightarrow\text{R} \qquad \dfrac{\dfrac{\phi \Rightarrow \phi}{\phi \Rightarrow T\ulcorner\phi\urcorner}\,\text{T2}}{\Rightarrow \phi \rightarrow T\ulcorner\phi\urcorner}\,\rightarrow\text{R}$$

□

Making use of the substitution function $\underset{\cdot}{s}$, it is easy to show that NTT derives all instances of the uniform T-schema as well. Despite many nice properties for a theory of truth, NTT falls short of other desiderata. In particular, it can be shown that NTT does not prove any new arithmetical sequents compared to PA and is therefore also not strong enough to prove PA's consistency.

Lemma 5.1.7. NTT is $\mathcal{L}_{\mathsf{PA}}$-conservative over PA.

If $\vdash_{NTT} \Rightarrow \phi$ with $\phi \in \mathcal{L}_{\mathsf{PA}}$, then $\vdash_{PA} \Rightarrow \phi$.

Proof. Assume for contradiction that there is a proof of a new arithmetical theorem, which is not provable in PA alone. Since it is not provable in PA, the proof needs to involve a T-rule at some point, yet the proved formula needs to be arithmetical. However, since Cut is blocked after making use of a T-rule, there is no rule which could eliminate the introduced T-formula. □

One might want to argue that the $\mathcal{L}_{PA}$-conservativity of the nontransitive theory is only a proof-theoretic weakness because it lacks elimination rules for T. In fact, it can be observed that classical theories closed under the schema $T\ulcorner\phi\urcorner \rightarrow \phi$ are very strong in (among other things) proving new arithmetical theorems (see e.g. (Cantini, 1990) and (Halbach, 2014)). However, this fact cannot be blamed on the absence of elimination rules for T, since they have been shown to be admissible in NTT. Thus adding them to the theory does not allow one to prove any new sequents, let alone new arithmetical ones not provable in PA. The following section explores ways of strengthening NTT in order to overcome this proof-theoretic weakness.

5.1.1 Strengthening NTT

NTT can be seen as a strengthening of a completely nontransitive theory of truth by allowing some Cuts back in given that its premises can be derived in PAT. The straightforward way to strengthen NTT is then to again consider stronger theories to derive the premises of our Cut rule. Thus our guiding question is *Which classical subtheories of* NT *can be made nontransitively transparent while remaining nontrivial?* We have shown that the strategy works for PAT and are now interested to see, for which extensions of PAT this holds as well.

We will study nontransitive theories based not only on PAT but on classical theories of truth, which extend PAT by some instances of the truth-rules:

Definition 5.1.2. Classical theory of truth

A classical theory of truth S is any consistent subtheory of NT, s.t. all rules but the compositional principles, T1 and T2 are unrestricted.

Note that PAT, by this definition, is a classical theory of truth as well. It is the special case in which all instances of the compositional principles, T1 and T2 are dropped. In the case of NTT above, we defined our nontransitive theory of truth by replacing the Cut rule of NT by one in which the premises must have been derived using only rules of PAT. We can generalise this restriction with the notion of a classical theory of truth S. In the following we define a class of nontransitive theories, which restricts the naive Cut rule to cases in which the Cut-premises have been derived using only instances of rules of S. Such a theory will be called Nontransitive Transparency over S:

Definition 5.1.3. Nontransitive Transparency over S

The theory NTT[S] (for 'Nontransitive Transparency over S') is defined by taking NT and replacing its rule of Cut by

$$\frac{\Gamma \Rightarrow \Delta, \phi \qquad \phi, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}\text{ SCut}$$

where in SCut, both its premises must be derived using only (instances of) rules of a classical theory of truth S. Note that the theory from the previous section NTT corresponds to $\mathsf{NTT[PAT]}$ in the new notation.

Theorem 5.1.8. Let S be a classical theory of truth. Then $\mathsf{NTT[S]}$ does not derive the empty sequent.

Proof. Analogous to the proof that NTT does not prove the empty sequent. □

Since S only needs to be classical and consistent, it makes sense to go for the strongest one. Most classical theories of truth in the literature take the compositional principles for granted and focus on restricting T1 and T2. So strong, classical theories amount to those with many instances of T1 and T2. So should we not simply consider the maximally consistent classical theory S with as many instances of T1 and T2 as possible? Closing this theory under the remaining instances of T1, T2 and the accordingly restricted Cut rule would then give us the strongest possible nontransitive theory possible.

Definition 5.1.4. Maximal consistency of S with respect to T1 and T2

A classical theory of truth S is maximally consistent with respect to T1 and T2 iff the addition of a new instance of T1 or T2 to S renders it inconsistent.

As is well-known due to (McGee, 1992), the quest for a maximally consistent classical theory of truth is in vain: He showed that there are infinitely many such theories and none of them are axiomatisable[1]. It is a short way from this result to show something very similar for nontransitively transparent theories of truth.

Definition 5.1.5. Maximally transitive transparency of S

A theory S is maximally transitively transparent iff a single addition of an instance of Cut to S renders it trivial.

[1]Strictly speaking McGee's paper discusses maximally consistent sets of instances of the T-schema. But due to the equivalence in NT between the T-rules and the T-schema, the theorem immediately transfers.

It is a short way from McGee's theorem to the analogous result for maximally transitive transparency. A nontransitively transparent extension of classical theory of truth S would be maximally transitively transparent iff S is maximally consistent with respect to T1 and T2. But since there is no such unique S and none of the candidates are axiomatisable, the same holds of maximally transitively transparent theories. But this only means that the search for the best logically possible restriction of Cut with respect to instances of T1 and T2 is futile. However, it still makes sense to look for subtheories of maximally consistent theories of truth, which are then extended by the missing T-rules and a restricted Cut-rule. In the following, we investigate such theories.

Proposition 5.1.9. For any S, NTT is a subtheory of NTT[S].

Corollary 5.1.9.1. For any S, NTT[S] is inconsistent and ω-inconsistent.

Proof. Immediate by the previous proposition, since NTT is both inconsistent and ω-inconsistent. □

Theorem 5.1.10. Nontrivial instances of NTT[S].

Let S be a classical theory of truth. Then NTT[S] is nontrivial.

Proof. Analogous to the proof of non-triviality for NTT. □

Lemma 5.1.11. NTT[S] is $\mathcal{L}_{PA}$-conservative over S.

Proof. Assume for contradiction that there are $\Gamma, \Delta \subseteq \mathcal{L}_{PA}$, s.t. NTT[S] proves $\Gamma \Rightarrow \Delta$ but S does not. Since the only rules which S compared to NTT[S] lacks are instances of T1 and T2, the proof of $\Gamma \Rightarrow \Delta$ must include at least one of these new instances of T1 or T2. But since the conclusion sequent $\Gamma \Rightarrow \Delta$ is arithmetical, the principal formula of this new T-inference must have been eliminated. The only rule, which could do so, is Cut. But since the proof includes an application of at least T1 or T2 not included in S, SCut cannot be applied. Thus there can be no such proof. □

5.1.2 A Strong Instance of NTT[S]

After investigating some basic properties of NTT[S] independently of some particular classical theory S, it is now time to look at some instances. One of the driving motivations to strengthen NTT to NTT[S] was to use the embedding of S in order to make our nontransitive theory proof-theoretically stronger. It should at least be able to prove the consistency of PA. The only thing we need to do is to find a classical theory of truth S, which already proves the consistency of PA and embed it into our nontransitive theory. Since NTT[S] proves everything S does, this makes sure that NTT[S] proves the consistency of PA as well.

But proving the consistency of PA is a relatively modest aim in terms of the proof-theoretic strength of formal theories of truth. Proving the consistency of a theory is inherently linked to 'how much' transfinite induction a theory is able to prove, i.e. what is the first ordinal the theory is not able to prove to be well-ordered. As shown originally by Gentzen (see Gentzen, 1969), transfinite induction up to ϵ_0 is necessary (and sufficient) in order to prove the consistency of PA. But ϵ_0 is a rather small proof-theoretic ordinal compared to some theories of truth, which prove much higher ordinals to be well-ordered.

Instead of going through various classical theories of truth of increasing strength, we simply show what the instance of NTT[S] looks like for the strongest theory of truth known so far. The strongest theory of truth is known as $\mathsf{UTB}(\mathsf{Z}_2)$ and was introduced by (Schindler, 2018). The idea behind the theory is both simple and genius. Instead of starting with some principles for T and hoping that the resulting theory is arithmetically stronger than PA, we start with a very strong arithmetical theory: Second-Order Arithmetic Z_2 formulated in a second-order language $\mathcal{L}_2$. Formulae of $\mathcal{L}_2$ are then translated into our first-order language of truth $\mathcal{L}_T$ and we add to PAT all instances of T1 and T2 where ϕ is a translation of a formula of $\mathcal{L}_2$ into $\mathcal{L}_T$.

In the following, we provide a partly simplified presentation of this powerful theory of truth. The language $\mathcal{L}_2$ of Z_2 is a two-sorted language, extending our arithmetical language $\mathcal{L}$ by second-order quantifiers and second-order variables $X, Y, ...$ as well as a binary relation symbol $\in$. The theory Z_2 can then be defined as the extension of PA formulated in $\mathcal{L}_2$ (including the formulation of the Induction scheme) by all instances of the comprehension scheme

$$\forall y_1, ..., y_m \forall Y_1, ..., Y_n \exists X \forall x (x \in X \leftrightarrow \phi(x, y_1, ..., y_m, Y_1, ..., Y_n))$$

Schindler then defines a primitive recursive translation function *, mapping formulae of $\mathcal{L}_2$ to formulae of $\mathcal{L}_T$ (see Schindler, 2018: p.454):

Definition 5.1.6. The translation function *

$x_i^* = x_{2i}$

$X_i^* = x_{2i+1}$

$\overline{0}^* = \overline{0}$

$(f(t_1, ..., t_2)^*) = f(t_1^*, ..., t_n^*)$

$(t_1 = t_2)^* = t_1^* = t_2^*$

$(\neg\phi)^* = \neg\phi^*$

$(\phi \rightarrow \psi)^* = \phi^* \rightarrow \psi^*$

$(t \in X_i^*) = T\underset{\cdot}{s}(x_{2i+1}, t^*)$

$(\forall x \phi)^* = \forall x_{2i} \phi^*$

$(\forall X_i \phi)^* = \forall x_{2i+1}(Fm_T^1(x_{2i+1}) \rightarrow \phi^*)$

Where Fm_T^1 applies to exactly those codes of sentences of $\mathcal{L}_T$, which contain exactly one free variable.

Definition 5.1.7. UTB(Z_2)

The theory UTB(Z_2) is defined by adding to PAT formuated in $\mathcal{L}_2$ all instances of the uniform T-schema

$$\forall x_1, ..., x_n (T \ulcorner \phi(\underset{\cdot}{x_1}, ..., \underset{\cdot}{x_n}) \urcorner \leftrightarrow \phi(x_1, ..., x_n))$$

where ϕ must be in the range of the translation function *.

We can easily give a sequent calculus for UTB(Z_2) by replacing the T1 and T2 of NT by

$$\frac{\phi, \Gamma \Rightarrow \Delta}{T\ulcorner\phi\urcorner, \Gamma \Rightarrow \Delta}\,\text{T1}^* \qquad \frac{\Gamma \Rightarrow \Delta, \phi}{\Gamma \Rightarrow \Delta, T\ulcorner\phi\urcorner}\,\text{T2}^*$$

where in both rules it must be the case that ϕ is in the range of *. It is then straightforward to embed $\mathsf{UTB}(\mathsf{Z}_2)$ into $\mathsf{NTT}[\mathsf{S}]$, obtaining $\mathsf{NTT}[\mathsf{UTB}(\mathsf{Z}_2)]$. We simply take $\mathsf{UTB}(\mathsf{Z}_2)$, add the instances of T1 and T2 s.t. their active formula is not in the range of * and demand of Cut that it is only applicable if both of its premises are derived using instances of rules of $\mathsf{UTB}(\mathsf{Z}_2)$.

Lemma 5.1.12. For any $\Gamma, \Delta \subset \mathcal{L}_2$, $\vdash_{UTB(Z_2)} \Gamma^* \Rightarrow \Delta^*$ iff $\vdash_{Z_2} \Gamma \Rightarrow \Delta$.

Proof. For a proof see (Schindler, 2018: p.473). □

Just how strong is $\mathsf{NTT}[\mathsf{UTB}(\mathsf{Z}_2)]$? Recall that some of the strongest theories of truth prove transfinite induction up to Γ_0. These theories are too weak to prove the comprehension scheme restricted to Π^1_1-formulae. In contrast to that, $\mathsf{UTB}(\mathsf{Z}_2)$ proves all (translated) instances of the comprehension scheme for Π^1_n – formulae for any $n \in \omega$. So $\mathsf{UTB}(\mathsf{Z}_2)$ and thus $\mathsf{NTT}[\mathsf{UTB}(\mathsf{Z}_2)]$ are very, very strong and much stronger than any other known classical or non-classical theory of truth. In fact, there is still no notation to capture the proof-theoretic ordinal of Z_2. The result of $\mathcal{L}_{\mathsf{PA}}$-conservativity for $\mathsf{NTT}[\mathsf{S}]$ over S then shows that $\mathsf{NTT}[\mathsf{UTB}(\mathsf{Z}_2)]$ does not prove any more arithmetical consequences than $\mathsf{UTB}(\mathsf{Z}_2)$ itself. So $\mathsf{NTT}[\mathsf{UTB}(\mathsf{Z}_2)]$ is as arithmetically strong as our strongest classical theory of truth.

The ability to 'absorb' the proof-theoretic strength of classical theories of truth, puts the nontransitive approach into a very powerful position compared to other non-classical approaches. Recently, proponents of classical theories of truth have argued against non-classical approaches by pointing out that they are much weaker than their classical counterparts compared to the arithmetical theorems they are able to prove. Where the classical theory KF (for Kripke-Feferman) proves transfinite induction up to ϵ_0, its non-classical cousin PKF (for Partial Kripke-Feferman) formulated in a paracomplete logic only proves transfinite induction up to ω^ω (see (Halbach and Nicolai, 2018) and (Halbach and Horsten, 2006)). (Picollo, 2018b) shows a similar result for the paraconsistent logic LP, in which classical inferences for some cases are gained back by using a recovery operator $\circ$.

The problem is that recapturing classical logical rules in these non-classical theories is hard. They cannot maintain classical logic for the full language $\mathcal{L}_T$, so what is typically done is to show that classical logic can be applied, whenever all premises are formulae of the base language $\mathcal{L}_{\mathsf{PA}}$. But proving the consistency of PA or the well-ordering of certain ordinals can of course not be done by working within the base language. We need the detour of using truth-theoretic principles and must thus prove sequents involving T-formulae. But here the classical principles can no longer be applied, which would be necessary to derive the new arithmetical theorems.

$\mathsf{NTT}[\mathsf{UTB}(\mathsf{Z}_2)]$ does not suffer from these issues of the non-classical approaches discussed above. We have shown that, given some classical theory of truth S, we can construct a nontransitive theory $\mathsf{NTT}[\mathsf{UTB}(\mathsf{Z}_2)]$, which has the same arithmetical consequences as S itself – and more truth-theoretic consequences, such as all instances of the (uniform) T-schema. Thus we conclude that our nontransitive approach fares much better with respect to recapturing the reasoning of classical theories of truth than other non-classical logics. $\mathsf{NTT}[\mathsf{UTB}(\mathsf{Z}_2)]$ has been shown to combine the strength of classical and non-classical theories of truth by being both proof-theoretically and truth-theoretically very strong.

5.1.3 An Alternative Formulation

So far we have expressed transitivity in our systems via Cut as the only elimination rule. But there are ways to make Cut admissible and thus absorb transitivity into other meta-inferences. We already saw this fact when we showed that Cut is equivalent to EME, a rule which allows one to eliminate instances of the law of excluded middle on the left hand side. Two other interesting elimination rules for negation and the conditional respectively are the following :

$$\frac{\Gamma \Rightarrow \Delta, \phi \qquad \Gamma \Rightarrow \Delta, \neg\phi}{\Gamma \Rightarrow \Delta}\ \neg\text{E} \qquad \frac{\Gamma \Rightarrow \Delta, \phi \qquad \Gamma \Rightarrow \Delta, \phi \to \psi}{\Gamma \Rightarrow \Delta, \psi}\ \to\text{E}$$

¬E and →E cannot be added to NTT without entailing triviality, as they make an unrestricted rule of Cut admissible:

Lemma 5.1.13. Cut is admissible in NTT + ¬E.

Proof. Assume that $\Gamma \Rightarrow \Delta, \phi$ and $\phi, \Gamma \Rightarrow \Delta$ are provable. Apply ¬R to the latter sequent to obtain $\Gamma \Rightarrow \Delta, \neg\phi$ and then apply ¬E together with the first sequent to conclude $\Gamma \Rightarrow \Delta$. □

The only problem to prove the admissibility of Cut when →E instead of ¬E is added, is that →E cannot conclude the empty sequent. This obstacle can be overcome by adding a falsity constant $\bot$, and let $\Rightarrow \bot$ be derived instead of the empty sequent. But nothing really hangs on this, since we now consider systems in which both rules are added instead of Cut[2]. Similarly, we can show that →E and ¬E are admissible in a theory with unrestricted Cut:

Lemma 5.1.14. →E and ¬E are admissible given Cut.

Proof. For the case of →E, assume that there are proofs of $\Gamma \Rightarrow \Delta, \phi$ and $\Gamma \Rightarrow \Delta, \phi \to \psi$. By the invertibility of →R, we get a proof of $\phi, \Gamma \Rightarrow \Delta, \psi$. Applying Cut gives us $\Gamma \Rightarrow \Delta, \psi$. To prove the admissibility of ¬E, assume proofs of $\Gamma \Rightarrow \Delta, \phi$ and $\Gamma \Rightarrow \Delta, \neg\phi$. An application of ¬L to the first premise gives $\neg\phi, \Gamma \Rightarrow \Delta$. Cut then concludes $\Gamma \Rightarrow \Delta$. □

Given the admissibility in both ways, we may conclude that →R and ¬E express just as much transitivity as Cut itself. This gives us the option of an alternative formulation of $\mathsf{NTT}[\mathsf{S}]$. Instead of restricting Cut, we drop the rule of Cut and replace it by restricted versions of →E and ¬E:

Definition 5.1.8. Noneliminable Transparency over S

The theory $\mathsf{NET}[\mathsf{S}]$ is defined by replacing the rule of Cut of $\mathsf{NTT}[\mathsf{S}]$ by

$$\frac{\Gamma \Rightarrow \Delta, \phi \qquad \Gamma \Rightarrow \Delta, \neg\phi}{\Gamma \Rightarrow \Delta}\ \text{S}\neg\text{E} \qquad\qquad \frac{\Gamma \Rightarrow \Delta, \phi \qquad \Gamma \Rightarrow \Delta, \phi \to \psi}{\Gamma \Rightarrow \Delta}\ \text{S}{\to}\text{E}$$

[2] Although the empty sequent cannot be derived by replacing Cut by →E, $\mathsf{NTT}[\mathsf{S}]$ still cannot be closed under →E unrestrictedly, since it renders the theory trivial. This is easily established via the Curry-paradox. Given premises $\Rightarrow T\ulcorner\kappa\urcorner$ and $T\ulcorner\kappa\urcorner \Rightarrow \phi$, we simply first introduce the conditional and then use →E to conclude $\Rightarrow \phi$ as also shown in (Ripley, 2015b)

where in both rules, both premises must have been derived using only (instances of) rules of S.

Corollary 5.1.14.1. For any S, $\mathsf{NTT}[\mathsf{S}]$ and $\mathsf{NET}[\mathsf{S}]$ are deductively equivalent.

Proof. Immediate by the admissibility of the elimination rules given Cut and vice versa. □

So we need not formulate the restriction of transitivity after certain T-inferences in terms of instances of Cut. We can also do so by considering the (meta-)rules of modus ponens and/or explosion, which are in a sense operational rules, since they are concerned with logical vocabulary.

5.2 An Indexed Calculus

So far our restriction on Cut was only expressed in the metalanguage. The fact tat the Cut premises must be derived in some classical theory S is neither expressible in our object language $\mathcal{L}_T$ nor in the language of our sequent calculus. Here we present a way of making this restriction more explicit by expressing it on the level of sequents. To this purpose, we introduce indices $i, j \in \{0, 1\}$, which are attached to the sequent arrow yielding indexed sequents $\Gamma \Rightarrow^i \Delta$. The intended reading of $\Gamma \Rightarrow^0 \Delta$ is that this sequent was derived using only instances of rules of S. $\Gamma \Rightarrow^1 \Delta$ then indicates that the sequent was derived using some instances of rules, which are not included in S itself.

Where arbitrary, we omit the indices for side formulae for the purpose of readability. In two-premise rules, if the two sequents have different indices i, j, the conclusion sequent has index $m(i, j)$. We use $m(i, j)$ to abbreviate $max\{i, j\}$ where max is the maximum function.

In the following we provide a full reformulation of $\mathsf{NTT}[\mathsf{S}]$ using indices and showing how the rules fare with indexed sequents. Given a classical theory of truth[3]

[3]Here we assume (as often before) that S includes the compositional principles unrestrictedly and only restricts T1 and T2.

S, an indexed formulation of $\mathsf{NTT[S]}$ – $\mathsf{NTT}[S]^i$ – is defined as the collection of the following rules:

$$\frac{}{\phi, \Gamma \Rightarrow^0 \Delta, \phi}\ Ref$$

$$\frac{}{s = t, \Gamma \Rightarrow^0 \Delta}\ {=}\text{F} \qquad \frac{}{\Gamma \Rightarrow^0 \Delta, s = t}\ {=}\text{T}$$

$$\frac{s = t, \Gamma \Rightarrow^i \Delta}{\Gamma \Rightarrow^i \Delta}\ {=E}$$

Where in $= T$, $s = t$ must hold an in =F, $s = t$ must not hold. As before, identity statements in these rules are limited to primitive recursive functions (for reasons of decidability).

$$\frac{\Gamma \Rightarrow^i \Delta, \phi}{\neg\phi, \Gamma \Rightarrow^i \Delta}\ \neg L \qquad \frac{\phi, \Gamma \Rightarrow^i \Delta}{\Gamma \Rightarrow^i \Delta, \neg\phi}\ \neg R$$

$$\frac{\Gamma \Rightarrow^i \Delta, \phi \qquad \psi, \Gamma' \Rightarrow^j \Delta'}{\phi \to \psi, \Gamma, \Gamma' \Rightarrow^{m(i,j)} \Delta, \Delta'}\ \to L \qquad \frac{\phi, \Gamma \Rightarrow^i \Delta, \psi}{\Gamma \Rightarrow^i \Delta, \phi \to \psi}\ \to R$$

$$\frac{\phi a, \Gamma \Rightarrow^i \Delta}{\forall x \phi x, \Gamma \Rightarrow^i \Delta}\ \forall L \quad \frac{\Gamma \Rightarrow^i \Delta, \phi y}{\Gamma \Rightarrow^i \Delta, \forall x \phi x}\ \forall R$$

$$\frac{\phi s, \Gamma \Rightarrow^i \Delta}{s = t, \phi t, \Gamma \Rightarrow^i \Delta}\ SubL \qquad \frac{\Gamma \Rightarrow^i \Delta, \phi s}{s = t, \Gamma \Rightarrow^i \Delta, \phi t}\ SubR$$

$$\frac{\Gamma \Rightarrow^0 \Delta, \phi \qquad \phi, \Gamma \Rightarrow^0 \Delta}{\Gamma \Rightarrow \Delta}\ Cut$$

$$\frac{}{0 = St, \Gamma \Rightarrow^0 \Delta}\ \text{Q1} \qquad \frac{}{Sr = St, \Gamma \Rightarrow^0 \Delta, r = t}\ \text{Q2}$$

$$\frac{}{\Gamma \Rightarrow^0 \Delta, r + 0 = r}\ \text{Q3} \qquad \frac{}{\Gamma \Rightarrow^0 \Delta, r + St = S(r + t)}\ \text{Q4}$$

$$\frac{}{\Gamma \Rightarrow^0 \Delta, r \times 0 = 0}\ \text{Q5} \qquad \frac{}{\Gamma \Rightarrow^0 \Delta, r \times St = (r \times t) + t}\ \text{Q6}$$

$$\frac{}{\forall x \neg(r = Sx), \Gamma \Rightarrow^0 \Delta, r = 0}\ \text{Q7}$$

$$\frac{\phi, \Gamma \Rightarrow^i \Delta}{T\ulcorner\phi\urcorner, \Gamma \Rightarrow^i \Delta}\ \text{T1[S]} \qquad \frac{\Gamma \Rightarrow^i \Delta, \phi}{\Gamma \Rightarrow^i \Delta, T\ulcorner\phi\urcorner}\ \text{T2[S]}$$

$$\frac{\phi, \Gamma \Rightarrow^i \Delta}{T\ulcorner\phi\urcorner, \Gamma \Rightarrow^1 \Delta}\ \text{T1} \qquad \frac{\Gamma \Rightarrow^i \Delta, \phi}{\Gamma \Rightarrow^1 \Delta, T\ulcorner\phi\urcorner}\ \text{T2}$$

Where in $T1[S]$ and $T2[S]$, it must be the case that S includes this instance of T1 or T2 respectively. T1 and T2 are only applicable if the theory does not include instances of T1[S] or T2[S] respectively for ϕ.

$$\frac{\Gamma, \neg Tt \Rightarrow^i \Delta}{\Gamma, Sent(t), T\underset{\cdot}{\neg} t \Rightarrow^i \Delta}\ \neg\mathrm{CL} \qquad \frac{\Gamma \Rightarrow^i \neg Tt, \Delta}{\Gamma, Sent(t) \Rightarrow^i T\underset{\cdot}{\neg} t, \Delta}\ \neg\mathrm{CR}$$

$$\frac{\Gamma, Tt \to Ts \Rightarrow^i \Delta}{\Gamma, Sent(t \underset{\cdot}{\to} s), T(t \underset{\cdot}{\to} s) \Rightarrow^i \Delta}\ \to\mathrm{CL} \qquad \frac{\Gamma \Rightarrow^i Tt \to Ts, \Delta}{\Gamma, Sent(t \underset{\cdot}{\to} s) \Rightarrow^i T(t \underset{\cdot}{\to} s), \Delta}\ \to\mathrm{CR}$$

$$\frac{\Gamma, \forall t Ts[t/v] \Rightarrow^i \Delta}{\Gamma, Sent(\underset{\cdot}{\forall} vs), T(\underset{\cdot}{\forall} vs) \Rightarrow^i \Delta}\ \forall\mathrm{CL} \qquad \frac{\Gamma \Rightarrow^i \forall t Ts[t/v], \Delta}{\Gamma, Sent(\underset{\cdot}{\forall} vs) \Rightarrow^i T(\underset{\cdot}{\forall} vs), \Delta}\ \forall\mathrm{CR}$$

It is easy to check that $\mathsf{NTT[S]}$ and $\mathsf{NTT[S]}^i$ are deductively equivalent (besides the indices of course).

5.2.1 Local And Global T-context

Formulated in terms of indices, a possible alternative restriction comes to mind: Instead of tracking which sequents were derived using instances of rules not included in S, one should track which formulae have been derived using such instances. Compare the following two applications of Cut (we assume that the shown applications of T1 and T2 are not included in S):

$$\frac{\dfrac{\psi, \Gamma \Rightarrow^0 \Delta, \phi}{T\ulcorner\psi\urcorner \Rightarrow^1 \Delta, \phi}\ \mathrm{T1} \qquad \phi, \Gamma \Rightarrow^0 \Delta}{T\ulcorner\psi\urcorner, \Gamma \Rightarrow^1 \Delta}\ \mathrm{Cut}$$

$$\frac{\dfrac{\Gamma \Rightarrow^0 \Delta, \phi}{\Gamma \Rightarrow^1 \Delta, T\ulcorner\phi\urcorner}\ \mathrm{T2} \qquad T\ulcorner\phi\urcorner, \Gamma \Rightarrow^0 \Delta}{\Gamma \Rightarrow^1 \Delta}\ \mathrm{Cut}$$

Both applications of Cut would be blocked in $\mathsf{NTT[S]}^i$, because the left Cut-premise is not derived using only instances of rules of S. However, there is an intuitively important difference between the two derivations. In the lower one, the

Cut-formula is derived using the problematic instance of a T-rule. In the upper derivation, the T-rule is only applied to some side formula. To distinguish what is happening in the two cases, we introduce the notions of a local and global T-context.

An inference is in a global T-context iff at least one of its premises was derived using T1 or T2. It is a in a local T-context iff at least one of its (occurrences of) active formulae was derived using T1 or T2. This distinction brings about the possibility of a less severe restriction on Cut: We restrict Cut iff its application is in a local T-context instead of whenever it is in a global T-context as before. In the following we introduce a calculus, which makes this idea precise and show that the local and global restrictions are deductively equivalent.

Locally Nontransitive Transparency

The distinction between a global and a local T-context gives two ways of restricting Cut – either if it is applied in a local or in a global T-context. $\mathsf{NTT}[\mathsf{S}]^i$ above reflects the restriction in a global T-context, which we will call the global restriction. Locally nontransitive transparency over S – $\mathsf{LNTT}[\mathsf{S}]^i$ – reflects the local restriction: Again, we make use of indices $i, j \in \{0, 1\}$. In this case, we do not sign sequents with indices but rather formulae, yielding ϕ^i. The intended interpretation of ϕ^0 is that ϕ was not derived using an instance of T1 or T2 not included in S. ϕ^1 respectively is read as ϕ being derived using such an instance. Where Γ is a set of formula, by Γ^i we mean that for all $\phi^j \in \Gamma^i$, it holds that $i = j$. As in the global case, only the instances of T-rules not included in S raise the index from 0 to 1, the rest of the rules determine the index of the conclusion depending on the indices of the premise(s). The theory $\mathsf{LNTT}[\mathsf{S}]^i$ is comprised of the following rules:

$$\frac{}{\phi^0, \Gamma^0 \Rightarrow \Delta^0, \phi^0}\ Ref$$

$$\frac{}{(s = t)^0, \Gamma^0 \Rightarrow \Delta^0}\ {=}\mathrm{F} \qquad \frac{}{\Gamma^0 \Rightarrow \Delta^0, (s = t)^0}\ {=}\mathrm{T}$$

$$\frac{(s = t)^0, \Gamma \Rightarrow^i \Delta}{\Gamma \Rightarrow^i \Delta}\ {=}\mathrm{E}$$

Where in =T and =E, $s = t$ must hold an in =F, $s = t$ must not hold and the identity statements are limited to primitive recursive functions.

$$\frac{\Gamma \Rightarrow \Delta, \phi^i}{\neg\phi^i, \Gamma \Rightarrow \Delta}\,\neg L \qquad \frac{\phi^i, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg\phi^i}\,\neg R$$

$$\frac{\Gamma \Rightarrow \Delta, \phi^i \qquad \psi^j, \Gamma \Rightarrow \Delta}{(\phi \to \psi)^{m(i,j)}, \Gamma \Rightarrow \Delta}\,\to L \qquad \frac{\phi^i, \Gamma \Rightarrow \Delta, \psi^j}{\Gamma \Rightarrow \Delta, (\phi \to \psi)^{m(i,j)}}\,\to R$$

$$\frac{\phi s^i, \Gamma \Rightarrow \Delta}{(\forall x \phi x)^i, \Gamma \Rightarrow \Delta}\,\forall L \qquad \frac{\Gamma \Rightarrow \Delta, \phi y^i}{\Gamma \Rightarrow \Delta, (\forall x \phi x)^i}\,\forall R$$

Where in ∀R, y must be an eigenvariable.

$$\frac{\phi s^i, \Gamma \Rightarrow \Delta}{(s = t)^0, \phi t^i, \Gamma \Rightarrow \Delta}\,SubL \qquad \frac{\Gamma \Rightarrow \Delta, \phi s^i}{(s = t)^0, \Gamma \Rightarrow \Delta, \phi t^i}\,SubR$$

$$\frac{\Gamma \Rightarrow \Delta, \phi^0 \qquad \phi^0, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}\,Cut$$

$$\frac{}{(0 = St)^0, \Gamma^0 \Rightarrow \Delta^0}\,\mathrm{Q1} \qquad \frac{}{(Sr = St)^0, \Gamma \Rightarrow \Delta^0, (r = t)^0}\,\mathrm{Q2}$$

$$\frac{}{\Gamma^0 \Rightarrow \Delta^0, (r + 0 = r)^0}\,\mathrm{Q3} \qquad \frac{}{\Gamma^0 \Rightarrow \Delta^0, (r + St = S(r + t))^0}\,\mathrm{Q4}$$

$$\frac{}{\Gamma^0 \Rightarrow \Delta^0, (r \times 0 = 0)^0}\,\mathrm{Q5} \qquad \frac{}{\Gamma^0 \Rightarrow \Delta^0, (r \times St = (r \times t) + t)^0}\,\mathrm{Q6}$$

$$\frac{}{(\forall x \neg(r = Sx))^0, \Gamma^0 \Rightarrow \Delta^0, (r = 0)^0}\,\mathrm{Q7}$$

$$\frac{\phi^i, \Gamma \Rightarrow \Delta}{T\ulcorner\phi\urcorner^i, \Gamma \Rightarrow \Delta}\,T1[S] \qquad \frac{\Gamma \Rightarrow \Delta, \phi^i}{\Gamma \Rightarrow \Delta, T\ulcorner\phi\urcorner^i}\,T2[S]$$

$$\frac{\phi^i, \Gamma \Rightarrow \Delta}{T\ulcorner\phi\urcorner^1, \Gamma \Rightarrow \Delta}\,T1 \qquad \frac{\Gamma \Rightarrow \Delta, \phi^i}{\Gamma \Rightarrow \Delta, T\ulcorner\phi\urcorner^1}\,T2$$

Where in $T1[S]$ and $T2[S]$, it must be the case that S includes this instance of T1 or T2 respectively. T1 and T2 are only applicable, if the theory does not include instances of T1[S] or T2[S] respectively for ϕ.

$$\frac{\Gamma, \neg Tt^i \Rightarrow \Delta}{\Gamma, Sent(t)^0, T\underset{\cdot}{\neg}t^i \Rightarrow \Delta}\,\neg\mathrm{CL} \qquad \frac{\Gamma \Rightarrow \neg Tt^i, \Delta}{\Gamma, Sent(t)^0 \Rightarrow T\underset{\cdot}{\neg}t^i, \Delta}\,\neg\mathrm{CR}$$

$$\frac{\Gamma, (Tt \to Ts)^i \Rightarrow \Delta}{\Gamma, Sent(t \dot{\to} s)^0, (T(t \dot{\to} s))^i \Rightarrow \Delta}\ {\scriptstyle \to\text{CL}}$$

$$\frac{\Gamma \Rightarrow (Tt \to Ts)^i, \Delta}{\Gamma, Sent(t \dot{\to} s)^0 \Rightarrow T(t \dot{\to} s)^i, \Delta}\ {\scriptstyle \to\text{CR}}$$

$$\frac{\Gamma, (\forall t Ts[t/v])^i \Rightarrow \Delta}{\Gamma, Sent(\dot{\forall} vs)^0, T(\dot{\forall} vs)^i \Rightarrow \Delta}\ {\scriptstyle \forall\text{CL}} \qquad \frac{\Gamma \Rightarrow (\forall t Ts[t/v])^i, \Delta}{\Gamma, Sent(\dot{\forall} vs)^0 \Rightarrow T(\dot{\forall} vs)^i, \Delta}\ {\scriptstyle \forall\text{CR}}$$

Theorem 5.2.1. $\vdash_{\mathsf{LNTT[S]}^i} \Gamma \Rightarrow \Delta$ iff $\vdash_{NTT[S]^i} \Gamma \Rightarrow \Delta$.

Proof. The right to left direction is trivial since any Cut in a local T-context is also in a global T-context (but not vice versa). The strategy for the left to right direction goes as follows. Assume for contradiction that there is some proof of a sequent $\Gamma \Rightarrow \Delta$ which is provable without Cut in a local T-context but not without Cut in a global T-context. We give a procedure to transform any such proof with a Cut in a global but not local T-context into one in which the Cut-inference is no longer in a global T-context.

Start with the last T-inference not included in S before the Cut-inference in question. Starting from this T-inference, go down the derivation to the first inference in which the principal formula of this T-inference or formulae in its trace are not active formulae. Then simply switch the last two inferences, i.e. the last inference in the trace of the principal T-formula and the first inference on the context of the T-inference. Repeat this process until the Cut-inference is pushed above the T-inference. For the purpose of illustration, consider the following example:

$$\dfrac{\dfrac{\dfrac{\psi s, \Gamma \Rightarrow \Delta, \phi}{\psi s, \Gamma \Rightarrow \Delta, T\ulcorner\phi\urcorner}\ {\scriptstyle \text{T2}}}{\psi s, \neg T\ulcorner\phi\urcorner, \Gamma \Rightarrow \Delta}\ {\scriptstyle \neg\text{L}}}{\forall x \psi x, \neg T\ulcorner\phi\urcorner, \Gamma \Rightarrow \Delta}\ {\scriptstyle \forall\text{L}} \quad \rhd \quad \dfrac{\dfrac{\dfrac{\psi s, \Gamma \Rightarrow \Delta, \phi}{\psi s, \Gamma \Rightarrow \Delta, T\ulcorner\phi\urcorner}\ {\scriptstyle \text{T2}}}{\forall x \psi x, \Gamma \Rightarrow \Delta, T\ulcorner\phi\urcorner}\ {\scriptstyle \forall\text{L}}}{\forall x \psi x, \neg T\ulcorner\phi\urcorner, \Gamma \Rightarrow \Delta}\ {\scriptstyle \neg\text{L}}$$

$$\rhd \quad \dfrac{\dfrac{\dfrac{\psi s, \Gamma \Rightarrow \Delta, \phi}{\forall x \psi x, \Gamma \Rightarrow \Delta, \phi}\ {\scriptstyle \forall\text{L}}}{\forall x \psi x, \Gamma \Rightarrow \Delta, T\ulcorner\phi\urcorner}\ {\scriptstyle \text{T2}}}{\forall x \psi x, \neg T\ulcorner\phi\urcorner, \Gamma \Rightarrow \Delta}\ {\scriptstyle \neg\text{L}}$$

After these transformations, we have pushed up the $\forall$L-inference on formulae in the context of the original T-inference above that T-inference. Iterating this procedure for all inferences on side formulae of the T-inference will push up the Cut in a global but not local T-context above the T-inference. But then it is no longer in a global T-context and thus becomes a proof of $\mathsf{NTT[S]}^i$ as well. Thus $\mathsf{NTT[S]}^i$ and $\mathsf{LNTT[S]}^i$ prove the same sequents (modulo indices of course). □

We conclude from this excursus that although the distinction between a local and a global T-context looks promising as a way of getting a more fine-grained restriction on Cut, this hope is in vain. Both restrictions yield the same theories in terms of provable sequents.

5.3 Transitive Non-Classical Approaches

The simplicity of the restriction and the strength of the resulting theory raises the question, whether it can be applied to other rules of NT. Especially the fact that it recovers strong classical theories completely makes it seem attractive for other non-classical approaches. As mentioned before, these approaches often struggle with recapturing classical logic. In the following we discuss the applicability of these ideas to well-known paracomplete, paraconsistent and noncontractive logics.

Paracomplete theories such as K_3 are non-classical theories, which take the lesson from the paradoxes to be that sentences such as the Liar are neither true nor false. Axiomatically, this amounts to dropping the law of excluded middle $\phi \vee \neg\phi$ in its general form and thus restricting any rule, which may derive it. In our setting of NT, this amounts to dropping the rule of ¬R and replacing it with some weaker negation rules but the details of these replacements do not matter here.

Paraconsistent theories such as LP accept the fact that the Liar and its negation are provable and so such sentences are both true and false. Since paraconsistent theories are (typically) closed under Cut, they cannot derive the principle of explosion $\phi \wedge \neg\phi \Rightarrow$. The derivation of this principle is blocked by giving up ¬L of NT and replacing it with some weaker principles for negation.

So both paracomplete and paraconsistent theorists might come up with the idea to play with the same recapture strategy of NTT[S]. However, we have already implicitly shown that no such project can work. For reconsider the derivation of the empty sequent via the Liar using the weak diagonal lemma and inversion:

$$\dfrac{\dfrac{\vdots}{\Rightarrow \lambda, T\ulcorner\lambda\urcorner}\;\dfrac{}{\Rightarrow T\ulcorner\lambda\urcorner}\text{T2} \qquad \dfrac{\dfrac{\vdots}{\lambda, T\ulcorner\lambda\urcorner \Rightarrow}}{T\ulcorner\lambda\urcorner \Rightarrow}\text{T1}}{\Rightarrow}\text{Cut}$$

The derivations leading up to $\Rightarrow \lambda, T\ulcorner\lambda\urcorner$ and $\lambda, T\ulcorner\lambda\urcorner \Rightarrow$ can be obtained in pure PAT. So the proofs up to this point must go through in the newly restricted theory as well. Further, non-classical approaches aim to be transparent theories of truth and so the following applications of T1 and T2 must go through as well. An application of Cut (also unrestricted in paracomplete and paraconsistent approaches) then shows that they still derive the empty sequent and are trivial by admissible Weakening (or the derivation for a Curry-sentence). The problem simply is that in deriving triviality, we do not need to use ¬L or ¬R s.t. their premises are derived using instances of rules not included in some classical theory of truth S. We only need to make use of ¬L or ¬R while we are still working in PAT.

However, it might be conjectured that every paradox p of NT contains an application of Contraction[4] s.t. its premise was derived using instances of rules not contained in S. Contraction expresses the property that repetitions of formulae in a sequent matter (see (Paoli, 2013) for different interpretations and motivations to restrict Contraction). For the most prominent proposal of noncontractive theories of truth see (Zardini, 2011) and (Grišin, 1982) for a noncontractive set theory. So far, we have dealt with sequents $\Gamma \Rightarrow \Delta$ where Γ, Δ are sets and these do not distinguish between multiple or a single occurrence of a formula. So in order to cash out Contraction, we momentarily interpret Γ, Δ as multisets $[\phi, \phi, \psi]$ of formulae, which are sensitive to the number of occurrences of the same formula in them. Where $\{\phi, \phi\} = \{\phi\}$, it holds for multisets that $[\phi, \phi] \neq [\phi]$. Contraction as a property of the consequence relation is then given by the following two rules:

$$\dfrac{\phi, \phi, \Gamma \Rightarrow \Delta}{\phi, \Gamma \Rightarrow \Delta}\text{CL} \qquad \dfrac{\Gamma \Rightarrow \Delta, \phi, \phi}{\Gamma \Rightarrow \Delta, \phi}\text{CR}$$

[4]By Contraction we always mean the structural rules shown below. In the literature, the scheme $(\phi \to (\phi \to \psi)) \to \phi \to \psi$ is sometimes called Contraction as well and plays a role in blocking the Curry-paradox.

Given $\to$R, we can show that there the conjecture above is false: There are paradoxes in which all applications of Contraction use premises derived fully in PAT. Obtain by weak diagonalisation a sentence $\alpha \leftrightarrow T\ulcorner\alpha\urcorner \to \neg T\ulcorner\alpha\urcorner$, which we call the 'Curry-Liar'. Informally, this sentence says 'If I am true, then I am not true'. By inversion we obtain sequents $\alpha, T\ulcorner\alpha\urcorner, T\ulcorner\alpha\urcorner \Rightarrow$, $T\ulcorner\alpha\urcorner \to \neg T\ulcorner\alpha\urcorner \Rightarrow \alpha$ and $\alpha \Rightarrow T\ulcorner\alpha\urcorner \to \neg T\ulcorner\alpha\urcorner$. The following derivations of $\Rightarrow \alpha$ and $\alpha \Rightarrow$ then go through, where the premise of the application of Contraction is derived in PAT:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\vdots}{\alpha, T\ulcorner\alpha\urcorner, T\ulcorner\alpha\urcorner \Rightarrow}}{T\ulcorner\alpha\urcorner, \alpha \Rightarrow}\,\text{CL}}{T\ulcorner\alpha\urcorner, T\ulcorner\alpha\urcorner \Rightarrow}\,\text{T1}}{\dfrac{T\ulcorner\alpha\urcorner \Rightarrow \neg T\ulcorner\alpha\urcorner}{\Rightarrow T\ulcorner\alpha\urcorner \to \neg T\ulcorner\alpha\urcorner}\,{\to}\text{R}}\,\neg\text{R} \qquad \dfrac{\vdots}{T\ulcorner\alpha\urcorner \to \neg T\ulcorner\alpha\urcorner \Rightarrow \alpha}}{\Rightarrow \alpha}\,\text{Cut}$$

$$\dfrac{\dfrac{\vdots}{\alpha \Rightarrow T\ulcorner\alpha\urcorner \to \neg T\ulcorner\alpha\urcorner} \qquad \dfrac{\dfrac{\vdots}{\Rightarrow T\ulcorner\alpha\urcorner} \quad \dfrac{\dfrac{\vdots}{\Rightarrow T\ulcorner\alpha\urcorner}}{\neg T\ulcorner\alpha\urcorner \Rightarrow}\,\neg\text{L}}{T\ulcorner\alpha\urcorner \to \neg T\ulcorner\alpha\urcorner \Rightarrow}\,{\to}\text{L}}{\alpha \Rightarrow}\,\text{Cut}$$

A Cut on $\Rightarrow \alpha$ and $\alpha \Rightarrow$ then gives the empty sequent. Thus such a restriction of Contraction cannot be successful in blocking paradoxes (and thus triviality) of NT. The Liar and the Curry-Liar show that the only rule which needs to be applied in a T-context in all paradoxes is Cut – at least relative to the setup of rules in NT. Of course, it might always be possible to reformulate ones sequent rules in order for them to become more fine grained in making different aspects of the consequence relation explicit. But we do not know of any such reformulations which would be fruitful in constructing a nontrivial theory of truth using the same idea as in NTT[S].

However, there might be a way for the noncontractivist around problematic derivations like the one via the Curry-Liar. Just as e.g. $\to$E makes Cut an admissible rule, there are formulations of operational rules which make at least some Contraction admissible. Compare the following two versions of the right conditional rule (of which the left one is included in NT):

$$\frac{\phi, \Gamma \Rightarrow \Delta, \psi}{\Gamma \Rightarrow \Delta, \phi \to \psi}\to\text{R} \qquad \frac{[\phi], \Gamma \Rightarrow \Delta, [\psi]}{\Gamma \Rightarrow \Delta, \phi \multimap \psi}\multimap\text{R}$$

Where the brackets $[\phi], \Gamma \Rightarrow \Delta, [\psi]$ indicate that either ϕ or ψ is taken as an active formula but not both. $\multimap$R is typically called the additive version of the right conditional rule, since (intuitively speaking) it only takes one of ϕ or ψ as a premise and 'adds' the other one. $\to$R is called the multiplicative version. With unrestricted rules of Contraction, it is easy to show that the two rules are interderivable. However, without Contraction, the following comparison shows that they come apart with respect to what is provable with them:

$$\frac{\phi, \Gamma \Rightarrow \Delta, \psi}{\Gamma \Rightarrow \Delta, \phi \to \psi}\to\text{R} \qquad \dfrac{\dfrac{\dfrac{\phi, \Gamma \Rightarrow \Delta, \psi}{\Gamma \Rightarrow \Delta, \psi, \phi \multimap \psi}\multimap\text{R}}{\Gamma \Rightarrow \Delta, \phi \multimap \psi, \phi \multimap \psi}\multimap\text{R}}{\Gamma \Rightarrow \Delta, \phi \multimap \psi}\text{CR}$$

If both immediate subformulae of $\phi \to \psi$ are present in the premise sequent, then we need Contraction to obtain a single occurrence of $\phi \multimap \psi$, whereas the same conclusion can be obtained via $\to$R directly without Contraction. This matters to paradoxes, because in the above derivation of the empty sequent via the Curry-Liar, we need to apply $\to$R in a T-context with both immediate subformulae of the conditional in the premise. So replacing $\to$R with $\multimap$R would make Contraction in a T-context necessary in the above derivation. Of course, this is no guarantee that there are no other paradoxes, in which no Contraction in a T-context needs to be applied despite having $\multimap$R. However, it is a necessary start for the noncontractivist and we do conjecture that with $\multimap$R replacing $\to$R and with Contraction explicit, every p of NT includes an application of Contraction in a T-context. But this turned out to be rather difficult to prove.

However, this gets us into the problem that structural rules are entangled with the operational rules: in the case of Contraction, $\to$R makes some cases of Contraction admissible. Thus restricting Contraction and replacing $\to$R by $\multimap$R, weakens the conditional. For example, we can no longer prove all instances of the T-schema, since some applications of Contraction in at least some T-contexts are blocked. Thus the following derivation no longer goes through for all instances of T1:

$$\dfrac{\dfrac{\dfrac{\dfrac{\phi \Rightarrow \phi}{T\ulcorner\phi\urcorner \Rightarrow \phi}\,\text{T1}}{\Rightarrow \phi, T\ulcorner\phi\urcorner \multimap \phi}\,{\multimap}\text{R}}{\Rightarrow T\ulcorner\phi\urcorner \multimap \phi, T\ulcorner\phi\urcorner \multimap \phi}\,{\multimap}\text{R}}{\Rightarrow T\ulcorner\phi\urcorner \multimap \phi}\,\text{CR}$$

Generally, we can only derive one direction of the T-schema for theorems $\Rightarrow T\ulcorner\phi\urcorner$ or $\Rightarrow \phi$ and anti-theorems $T\ulcorner\phi\urcorner \Rightarrow$ or $\phi \Rightarrow$. For in these cases there is only one of the immediate subformulae of the conditional present. So although a (working) noncontractive solution would be transparent due to the unrestricted T-rules, its conditional is too weak to express this fact in the object language $\mathcal{L}_T$ as it is not able to prove all instances of the (uniform) T-schema.

There is also another important issue for the noncontractivist who tries to restrict Contraction in T-contexts. It turns out that a noncontractive approach cannot be strengthened by embedding any consistent theory of truth S as was possible in $\mathsf{NTT}[\mathsf{S}]$. For suppose that S is consistent and closed under $T\ulcorner\phi\urcorner \to \phi$ as some of the strongest theories of truth like KF of (Reinhardt, 1986; Feferman, 1991) or VF of (Cantini, 1990) are. This amounts to a closure under T1 in our sequent calculus setting. It is then easy to show that S derives $\Rightarrow \lambda$ and $T\ulcorner\lambda\urcorner \Rightarrow$ for our standard Liar sentence λ. A noncontractive theory over S blocking Contraction-inferences after applications of T2 not contained in S derives the empty sequent and is thus trivial:

$$\dfrac{\dfrac{\begin{array}{c}\vdots\\ \Rightarrow \lambda\end{array}}{\Rightarrow T\ulcorner\lambda\urcorner}\,\text{T2} \qquad \begin{array}{c}\vdots\\ T\ulcorner\lambda\urcorner \Rightarrow\end{array}}{\Rightarrow}\,\text{Cut}$$

So some classical theories of truth (despite their consistency) already contain ‘too much’ Contraction, s.t. an embedding into a noncontractive solution is not possible. Our nontransitive approach $\mathsf{NTT}[\mathsf{S}]$ has neither of the two problems discussed above.

Another difference between applying the idea of restricting rules in a T-context to the noncontractive or the nontransitive approach is the behaviour with respect to the distinction between a local and a global T-context. In the previous section we have shown that whether we restrict Cut locally or globally makes no difference with respect to which sequents are provable. This is not the case in the noncontractive approach. This is because due to the unrestricted Cut-rule in the noncontractive approach, we can always change active formulae of applications of Contraction with index 1 for occurrences with index 0 as follows:

$$\dfrac{\dfrac{\begin{matrix}\vdots \\ \Gamma \Rightarrow \Delta, \phi^0, \phi^1\end{matrix} \qquad \phi^0 \Rightarrow \phi^0}{\Gamma \Rightarrow \Delta, \phi^0, \phi^0}\,\text{Cut}}{\Gamma \Rightarrow \Delta, \phi^0}\,\text{CR}$$

Since ϕ is arbitrary, this works in the case of paradoxes as well, allowing derivations of the empty sequent and triviality in the noncontractive approach when Contraction is restricted only locally.

Chapter 6

Assessing Theories Of Truth

This chapter is concerned with the assessment of theories of truth, discussing various criteria a theory of truth should fulfil. This allows us to assess our nontransitive theory NTT[S] with respect to these criteria and to compare it with other rival accounts. To do so, the chapter is divided into three parts: The first part discusses what role the truth predicate plays or should play in formal theories of truth. This gives us a list of functional roles the truth predicate should be able to play in a formal theory of truth and we can judge a theory of truth as adequate or inadequate with respect to these functional roles. The second part focuses on theoretical virtues of theories of truth by giving a systematic overview over the (still young and unexplored) work in the literature. This list is extended by further virtues taken from the literature on virtues in science and the philosophy of science in the third section.

After the discussion of each set of assessment criteria for a theory of truth, these criteria are applied to our nontransitive theory. Overall, we argue that NTT[S] fares very well with respect to many of the discussed virtues, in fact much better than most of the other theories of truth on the market. In particular, we argue that NTT[S] is able to improve upon classical and other non-classical theories of truth.

6.1 Functional Desiderata

A good theory of truth is one in which truth is able to fulfil certain functional roles. So in order to determine how well a truth theory behaves and to be able to compare it with other theories of truth in these respects, we first need to clarify the functional roles our truth predicate T plays. We distinguish two major groups of functional roles of the truth predicate: substitutional and quantificational. In the following, we discuss and explain these two groups and show how functional roles identified in the literature fit into this classification.

6.1.1 Substitutional Roles

There are different substitutional roles of the truth predicate as we will see in more detail below, but they all have to do with the idea that at least in certain contexts, ϕ can be replaced by $T\ulcorner\phi\urcorner$ and vice versa. We first describe some examples from natural language where this replacement plays a role before we discuss ways in which this idea can be made precise.

(Picollo and Schindler, 2019) describe what they call the *redundant* and *blind* uses of the truth predicate, which are both forms of the substitutional role of T in natural language in our sense. The *redundant* use is exemplified in cases in which the truth predicate is simply applied to a single sentence. For example, one may utter the sentence "Snow is white' is true'. This use of the truth predicate, however, is redundant. For we would convey the exact same message or content by simply uttering the sentence 'Snow is white' without making use of the truth predicate.

When applying the predicate to names of sentences one might not know explicitly, we are using the truth predicate in the *blind* sense. They give the example of someone using the sentence "Goldbach's conjecture' is true'. This use is different from the previous, redundant, one because, dependent on the knowledge of the speaker, it cannot be so easily dismissed with as in the previous case. For if the individual does not know what Goldbach's conjecture is, or cannot express it for some reason, she would not be able to express her acceptance of the conjecture without making use of the truth predicate.

Both uses of the truth predicate, the redundant and blind one, can be formalised in terms of some replacibility of ϕ and $T\ulcorner\phi\urcorner$ in sequents. Formally, the possibility to replace ϕ by $T\ulcorner\phi\urcorner$ and vice versa comes in different degrees. The weakest degree comes in the form of the rules of NEC and CONEC:

$$\frac{\Rightarrow \phi}{\Rightarrow T\ulcorner\phi\urcorner}\ \text{NEC} \qquad \frac{\phi \Rightarrow}{T\ulcorner\phi\urcorner \Rightarrow}\ \text{CONEC}$$

In the case of NEC and CONEC, the substitutability is very restricted as it only applies to formulae in sequents with no side formulae.

The redundant and blind use are contained in the idea that the truth predicate ought to be transparent. As mentioned before, the truth predicate of a theory is transparent iff: it proves $\phi, \Gamma \Rightarrow \Delta$ iff it proves $T\ulcorner\phi\urcorner, \Gamma \Rightarrow \Delta$ and it proves $\Gamma \Rightarrow \Delta, \phi$ iff it proves $\Gamma \Rightarrow \Delta, T\ulcorner\phi\urcorner$. Many logicians who take a nonclassical approach to theories of truth advocate transparency as a key feature of such theories, including proponents of nontransitive theories such as Ripley (see e.g. Ripley, 2013a). This feature is clearly fulfilled by $\mathsf{NTT}[\mathsf{S}]$ given that the theory in question is closed under the truth rules T1-T4. $\mathsf{NTT}[\mathsf{S}]$ contains T1 and T2 unrestrictedly, T3 and T4 are admissible as shown in 2.2.4.

(Field, 2008) emphasises that the substitutional roles of the truth predicate are not fulfilled by a merely transparent truth predicate. Rather than being merely transparent, the truth predicate must be *intersubstitutable*. Whereas transparency allows us to substitute ϕ for $T\ulcorner\phi\urcorner$ and vice versa in sequents, intersubstitutivity goes further than that. For the truth predicate to be intersubstitutable, we need to be able to substitute ϕ for $T\ulcorner\phi\urcorner$ in formulae. So for example, we need that $\Gamma \Rightarrow \Delta, \phi \rightarrow \psi$ is derivable iff $\Gamma \Rightarrow \Delta, T\ulcorner\phi\urcorner \rightarrow \psi$ is. The same goes for other logical constants such as negation.

To motivate such a desiderata, suppose you are a politician listening to an elaborate speech of a climate scientist talking about the causes and consequences of global warming. You are impressed by the speech and in particular convinced by the large number of arguments and facts presented to you. It is also clear to you that the consequence of these facts is to radically change the political course of your party. You would like to express the linkage between the scientist's speech and the political change in terms of a conditional. But instead of listing all the facts $f_1, f_2, ..., f_n$, which all seem relevant for your decision, you use the truth predicate and simply say

> If everything the scientist said in his speech is true, we must change our political course.

Of course you rely on the fact that your uttered sentence is equivalent to 'If $f_1 \wedge f_2 ... \wedge f_n$, then we must change our political course'. But this equivalence demands that the facts and the result of applying the truth predicate to them (given some quotation device) are intersubstitutable. What we need is not just that facts $\Rightarrow \phi$ and $\Rightarrow T\ulcorner\phi\urcorner$ are interderivable but also that ϕ and $T\ulcorner\phi\urcorner$ are intersubstitutable in hypothetical contexts – as in the example of an antecedent of a conditional. The examples above speak only of the intersubstitutivity of the truth predicate in propositional connectives, i.e. where the main connective of the formula in question is a logical constant of propositional logic.

Definition 6.1.1. Intersubstitutivity of T

A theory of truth S has an intersubstitutable truth predicate iff S contains the following rules or they are admissible in S:

$$\frac{\neg\phi, \Gamma \Rightarrow \Delta}{\neg T\ulcorner\phi\urcorner, \Gamma \Rightarrow \Delta}\ \neg\text{TL} \qquad \frac{\Gamma \Rightarrow \Delta, \neg\phi}{\Gamma \Rightarrow \Delta, \neg T\ulcorner\phi\urcorner}\ \neg\text{TR}$$

$$\frac{\phi \rightarrow \psi, \Gamma \Rightarrow \Delta}{\phi^\dagger \rightarrow \psi^\dagger, \Gamma \Rightarrow \Delta}\ \rightarrow\text{TL} \qquad \frac{\Gamma \Rightarrow \Delta, \phi \rightarrow \psi}{\Gamma \Rightarrow \Delta, \phi^\dagger \rightarrow \psi^\dagger}\ \rightarrow\text{TR}$$

where † in →TL and →TR indicates that either ϕ or ψ has been replaced by either $T\ulcorner\phi\urcorner$ or $T\ulcorner\psi\urcorner$ respectively (or both). The double lines indicate that the rule can be applied both ways – from bottom to top and top to bottom.

That the truth predicate of $\mathsf{NTT[S]}$ is intersubstitutable in this sense can be shown as follows:

Lemma 6.1.1. The truth predicate of $\mathsf{NTT[S]}$ is intersubstitutable.

Proof. The admissibility of all four rules can be shown with the help of the invertibility of our rules for $\neg$ and $\rightarrow$. Consider first only the top to bottom directions. For the admissibility of $\neg$TL, assume that $\neg\phi, \Gamma \Rightarrow \Delta$ is provable. By the invertibility of $\neg$L, it follows that $\Gamma \Rightarrow \Delta, \phi$ is provable. An application of T2, followed by an application of $\neg$L yields $\neg T\ulcorner\phi\urcorner, \Gamma \Rightarrow \Delta$. The case for $\neg$TR works analogously. For the case of $\rightarrow$TL, assume that $\phi \rightarrow \psi, \Gamma \Rightarrow \Delta$ is provable. By the invertibility of $\rightarrow$L, it follows that $\Gamma \Rightarrow \Delta, \phi$ and $\psi, \Gamma \Rightarrow \Delta$ are provable. An application of T1 or T2 yields either $\Gamma \Rightarrow \Delta, T\ulcorner\phi\urcorner$ or $T\ulcorner\psi\urcorner, \Gamma \Rightarrow \Delta$ (depending on which subformula should be replaced by the T-formula). An application of $\rightarrow$L then yields $\phi^\dagger \rightarrow \psi^\dagger, \Gamma \Rightarrow \Delta$. The case of $\rightarrow$TR works analogously by inversion. The bottom to top directions work the same way. First use inversion, then the admissibility of T3 or T4 to eliminate the truth predicate and then reintroduce the constant. □

Thus NTT[S] fulfils the desideratum of having a propositionally intersubstitutable truth predicate and fulfils all discussed substitutional roles.

6.1.2 Quantificational Roles

Where the substitutional roles had to do with the possibility to switch from ϕ to $T\ulcorner\phi\urcorner$ in different contexts, the quantificational uses are concerned with the utility of using quantifiers in formula containing the truth predicate. For the case of quantifying over finitely many objects while using the truth predicate, consider the example

> Everything that Carnap said is true.

Since Carnap (unfortunately) lived only a finite amount of time with finite resources, he has only uttered finitely many sentences. So our example expresses a finite generalization. Here we must note two things regarding the dispensability of the truth predicate when using it for finite generalizations. First, the generalization above could have been expressed using sentential quantification. If we allow for variables p ranging over sentences (or propositions), giving us quantification of the form $\forall p \phi$ and letting Cp express that 'Carnap said p', then the sentence above can be formalised as $\forall p(Cp \rightarrow p)$. Second, the truth predicate can be made redundant in the sense of the redundant use discussed above. Let $p_1, p_2, ..., p_n$ be the list of sentences (or propositions expressed thereby) uttered by Carnap. Then the above sentence can be rewritten as the conjunction of the list of sentences, so $p_1 \wedge p_2, ... \wedge p_n$.

Things look a bit different in the case of infinite generalizations. (Picollo and Schindler, 2019) give the example

> Every theorem of PA is true.

The difference to the sentences uttered by Carnap during his lifetime is of course that there are infinitely many theorems of PA. Still, the truth predicate is redundant in the sense that it could be replaced via sentential quantification: $\forall p(Prov_{\mathsf{PA}} p \rightarrow p)$. However, it cannot be so easily replaced by making use of conjunctions. Since there are infinitely many theorems of PA, the conjunction would have to have infinitely many conjuncts. Of course we can build formal theories in which such infinite conjunctions are allowed. However, the point here is that natural language is not expressive enough to include such conjunctions. Human speakers are not able to make use of such conjunctions. So outside the scope of such artificial languages, we cannot dispense of sentential quantification or the truth predicate (or maybe some other finite quantificational device) in order to express such sentences as the one about the truth of the theorems of PA.

Famously, Quine already saw this crucial difference between finite and infinite generalizations when it comes to the use of the truth predicate:

> We may affirm the single sentence by just uttering it, unaided by quotation or by the truth predicate; but if we want to affirm some infinite lot of sentences that we can demarcate only by talking about the sentences, then the truth predicate has its use. (Quine, 1970: p.12)

Expressing finite or infinite generalizations as the ones above can be done using sentential (propositional) quantification or the truth predicate (or possibly other linguistic devices). So the truth predicate allows us to emulate this sentential quantification without introducing it explicitly into our language. There are also other cases in which the truth predicate helps us with issues of quantification, namely regarding quantification into predicate position. (Picollo and Schindler, 2019) use the following example to make this point clear:

Tom is mortal.

Based on this sentence, we may want to infer that 'There is a property X, s.t. Tom exhibits X' and by doing so, we would quantify into predicate position. This is simply to say that we introduce variables ranging over predicates and bind these variables by using appropriate quantifiers. First order logic does not allow this, since variables only range over objects, not over predicates. One easy way out, similar to using sentential quantification in the examples above, is to make use of second-order logic. Such logical systems introduce new variables ranging over predicates which can then be bound by second-order quantifiers. Again, analogous to the case of sentential quantification, the truth predicate allows us to mimick this second-order quantification. Consider the sentence

'Tom is mortal' is true.

What happens by going from 'Tom is mortal' to "Tom is mortal' is true' is not only the introduction of the truth predicate but also, crucially, the syntactic move from a sentence to an object. By introducing the quotation device on the sentence 'Tom is mortal' we have syntactically transformed it into an object to which the truth predicate applies – and we can quantify over objects in our first-order languages and systems. To see how this works, consider the operation of concatenation of objects by which we simply mean binding two objects together. So the result of applying concatenation to Tom and 'is mortal' is 'Tom is mortal'. Considering the sentence 'The result of applying concatenation to Tom and 'is mortal' is true', quantification into predicate position as shown above can then be done as follows using the truth predicate:

> There is an x, s.t. the result of applying concatenation to 'Tom' and x is true.

So how does NTT[S] fare with respect to the quantificational role of the truth predicate, how much second-order quantification is it able to emulate? To answer this question, it makes sense to assess a theory of truth over PA in this respect by comparing it to the result of adding second-order quantification to PA instead of the truth predicate. This way we can see, just how much of the functional role of quantification the truth predicate in some formal theory of truth is able to play (at least relative to the case of Peano Arithmetic). As discussed before, the result of extending PA by second-order quantifiers and all instances of the comprehension scheme is full second-order arithmetic Z_2. Further, we have shown that there is an instance of NTT[S], NTT[UTB(Z_2)], which proves all translations of sequents provable in Z_2. So there is a clear sense in which NTT[UTB(Z_2)] fulfils the quantificational role of the truth predicate.

However, there are two restrictions which apply to the emulation of second-order quantification via the truth predicate. First, Picollo & Schindler point out that

> Truth theories can handle quantification into predicate position by talking about the concatenation of certain terms with certain formulas being true, as indicated above. But this means that truth theories can only simulate predicate quantification for those languages that contain a name for each individual in the domain of the first-order quantifiers – thus, the domain needs to be countable. (Picollo and Schindler, 2019: p.334)

This first concession need not really bother us. At least for now, it is hard to see what motivation there would be to add a truth predicate to an uncountable language. At least if our motivations to study truth, roughly, come from reflections on natural language, this does not limit our project at all. Further, if there is a need or an interest in emulating second-order quantification via a predicate, Picollo and Schindler go on to show that this limitation can be overcome by switching from a truth predicate to a satisfaction predicate. Second, Schindler concedes in a different paper that

> While second-order languages allows us to generalize over any predicate of the second-order language, the truth predicate of $\mathsf{UTB}(\mathsf{Z}_2)$ does not allow us to generalize over all predicates of the language of truth, but only over those predicates that can be translated back into the second-order formalism. (Schindler, 2018: p.473)

This remark does not really tell us something new about the kind or range of emulating second-order quantification. Of course, using the translation from $\mathcal{L}_2$ into $\mathcal{L}_T$ is not the same as adding second-order quantifiers to $\mathcal{L}_T$. But such a limitation is not a shortcoming of the translation function and does not reveal any problems for the project overall. After all, we already knew real second-order logic to be too powerful (as exemplified by its lack of compactness) in order to be fully mimicked by the truth predicate.

In the previous two sections we have identified two major categories of functional roles of the truth predicate: Substitutional roles and quantificational roles. Our assessment has shown that our nontransitive approach fulfils all of these roles. Choosing a suitable classical theory S such as $\mathsf{UTB}(\mathsf{Z}_2)$ allows $\mathsf{NTT}[\mathsf{UTB}(\mathsf{Z}_2)]$ to mimick second-order quantification over arithmetical predicates. Thus it fulfils the quantificational function of the truth predicate. Further, we have seen that the nontransitive approach fulfils the substitutional functions. By unrestricted T1 and T2, ϕ is derivable whenever $T\ulcorner\phi\urcorner$ is and the theory is able to non-trivially deal with substitution rules T1† and T2† for the truth predicate.

This already puts our approach in a very favourable position compared to other theories of truth: First, concentrating on only either the substitutional or the quantificational role, $\mathsf{NTT}[\mathsf{UTB}(\mathsf{Z}_2)]$ already fares much better than other theories of truth. This is especially clear with respect to the quantificational role. We do not know of any non-classical theories of truth which are able to mimick 'as much' second-order quantification as $\mathsf{NTT}[\mathsf{UTB}(\mathsf{Z}_2)]$ is able to do. In fact, non-classical theories of truth are often much weaker from a proof-theoretic perspective than their classical cousins. The classical axiomatisation of the Kripke fixed-point construction KF proves transfinite induction for the language of truth up to Γ_0, whereas the non-classical axiomatisation PKF only proves transfinite induction for $\mathcal{L}_T$ up to ω^ω (see Halbach and Horsten, 2006). So already considering just one of the functional roles shows that our nontransitive approach is superior compared to other non-classical approaches.

Second, NTT[S] is superior to classical theories of truth because it is not only able to fulfil the quantificational roles but also the substitutional roles *at the same time*. Although at least some classical theories fulfil the quantificational roles, their truth predicate does not allow for an unrestricted redundant use, let alone full intersubstitutivity. This is because the classical approach, by maintaining all classical rules, must by Tarski's theorem put some restrictions on T1 and T2. Thus the full redundant use and intersubstitutivity are lost. This is not the case with NTT[S]: Here we have a truth predicate which plays both roles – the substitutional and the quantificational one.

6.2 Theoretical Virtues

In the previous sections, we have explored what functional roles we should expect a truth predicate to play in a formal theory of truth over PA and determined that our nontransitive approach is able to fulfil this demand. However, when discussing and comparing theories of truth, such functional roles often play a subordinate role (we do not think this is the right approach, but that is a different matter for now). Rather, many authors in the literature on truth theories focus on *theoretical virtues* of such theories. This section gives a systematic overview over such theoretical virtues as they have been discussed in the literature and argues that our nontransitive approach fulfils many of them.

One might object from the outset against considering theoretical virtues in assessing theories of truth as such virtues are rather applied in assessing theories in natural science. After all, their subject matter and methodology are vastly different. One might even go so far to say that the enterprise of a formal theory of truth is a priori, whereas the one of natural sciences is a posteriori. So since the two are so vastly different, one cannot transfer the methodology of the one to the other. Drawing such a categorical wedge between formal theories (such as those of truth) and empirical theories has a long tradition. However, this tradition has been questioned quite extensively in the recent literature – for good reasons as we think.

Defenders of the classical idea that there is a categorical difference between formal theories and empirical theories take it that this difference has clear consequences for the possibility of revising one's theory. Where it is possible and often necessary to change ones empirical theory in the light of new data, such a change is in principle not possible for formal theories such as logic or mathematics. The principles of such theories (and thus their consequences) are eternal and analytic, they cannot be changed. One of the most prominent defenders of the alternative view that even logic can in principle be changed is Quine.

Recently, such views have gained new interest, especially in the work of Hjortland and Williamson (see e.g. Hjortland, 2017; Williamson, 2017, 2018). They argue that empirical and formal theories are not radically different when it comes to their methodology, motivation and possibility of revision. In all these cases, one should follow an abductive methodology. This idea that there is nothing special about formal theories such as logic is called anti-exceptionalism. In the following, we assume a certain amount of anti-exceptionalism without arguing for it explicitly, at least to a degree which allows us to apply abductive criteria such as theoretical virtues to the theory choice when it comes to picking a formal theory of truth. In the following two subsections we first explore theoretical virtues from the literature on theories of truth and then virtues from the literature in the philosophy of science.

6.2.1 Virtues For Theories Of Truth

The current locus classicus of theoretical virtues in theories of truth is (Leitgeb, 2007). Leitgeb gives the following list of virtues:

1. Truth is a predicate applied to objects given by some syntax theory which is part of the theory of truth.

2. Adding a theory of truth to some base theory (mathematical or empirical), the resulting truth theory should prove the consistency of the base theory.

3. No type restrictions should apply to T.

4. Instances of the T-schema should be derivable for all sentences of the language.

5. The truth predicate should adhere to compositional principles regarding the connectives and quantifiers.

6. The theory of truth should have an ω-model.

7. The outer and inner logic should be the same.

8. The outer logic should be classical logic.

Before we comment on this list of virtues, a few explanations are necessary. What exactly amounts to typing is up to some philosophical debate, but the term is widely used in two senses in the literature on theories of truth. A truth predicate can be typed syntactically if there are restrictions with respect to what counts as a well formed formula in the formal language count as well-formed. For example, we could define a sublanguage of $\mathcal{L}_T$ in which the truth predicate can only be applied to sentences themselves not including the truth predicate. In the case of axiomatic theories of truth, one sometimes speaks of typed theories, if the T-schema (or respective rules for T) are restricted in the same sense. The rules for T can only be applied to formulae themselves not containing the truth predicate.

What lies behind the idea of the compositionality of truth is that the truth of a complex sentence should be equivalent to the truth of its components. The truth of a conjunction should hold iff all conjuncts are true and a universal formula $\forall x \phi x$ should be true iff all instances of ϕx are. Compositionality can be expressed differently in axiomatic theories of truth. A theory of truth might be able to prove all instances of the schema $T \dot{\neg} \ulcorner \phi \urcorner \leftrightarrow \neg T \ulcorner \phi \urcorner$ or it might be able to prove the sentence $\forall x(Sentx \rightarrow (T \dot{\neg} x \rightarrow \neg Tx))$ or contain this sentence as an axiom as compositional theories of truth typically do. In NT, the compositionality is expressed in terms of the compositional principles in the form of sequent rules.

The criterion regarding the ω-model really has two components: consistency and ω-consistency. Consistency (assuming classical logic) is an immediate requirement, since otherwise the theory would be trivial. Regarding ω-consistency, Leitgeb argues that a theory of truth would otherwise rule out the intended interpretation of symbols of the arithmetical base theory and mess up the original ontological commitments:

> A theory of truth should not exclude this standard interpretation, for otherwise the theory could not be understood as speaking about the very objects that it was designed to refer to. Put differently: a theory of truth does not only have to be consistent (of course it has to be!), it also should not mess up its intended ontological commitments. (Leitgeb, 2007: p.281).

The distinction between inner and outer logic has to do with the scope of the truth predicate. The idea is to distinguish which logical principles apply to sentences ϕ and $T\ulcorner\phi\urcorner$. In the case of ϕ (assuming it is not of the form Ts) we speak of the outer and in the case of ϕ under the scope of the truth predicate we speak of the inner logic. As Leitgeb points out, there are theories of truth, which prove all instances of $\phi \vee \neg\phi$ but also prove $\neg T(\lambda \vee \neg\lambda)$, i.e. that exluded middle for the Liar sentence λ is not true. (Horsten and Halbach, 2015) point out that this criterion can be reduced to the demand that all instances of the T-schema should be provable and that the outer logic should be classical.

Before we assess our nontransitive theory of truth with respect to these virtues, we extend the list by some further considerations and try to give a systematic account of all relevant virtues. (Sheard, 2013) gives another list of desiderata for a theory of truth:

1. Truth is an objective concept, which applies independently of such things as desires or thoughts.
2. Provability preserves truth, i.e. the inference from $\vdash \phi$ to $\vdash T\ulcorner\phi\urcorner$ is available unrestrictedly.
3. Constructions to determine the extension of T should be simple, choices of axioms should follow some principle and not be ad hoc.
4. It is fine for different theories of truth to serve different purposes.
5. A theory of truth should be closed under infinitary closure, i.e. $\forall x(T\ulcorner\phi \dot{x}\urcorner \rightarrow T\ulcorner\phi x\urcorner)$.

Some desiderata turn out to be special cases of items on Leitgeb's list, others expand it. The first two and the last desiderata are special cases of things we have already discussed above: That truth is an objective concept is guaranteed by our setup as a formal theory of truth with a suitable arithmetical base theory. The rule of Necessitation, to go from $\vdash \phi$ to $\vdash T\ulcorner\phi\urcorner$, follows from an unrestricted T-schema and the infinitary closure is part of the compositionality criterion mentioned before. Desiderata three and four about simplicity, non-adhocness and the fact that different theories can serve different purposes are not included in the virtues discussed so far. We will discuss them in detail in the next section on virtues from natural science.

The last paper we will consider on desiderata for theories of truth is (Horsten and Halbach, 2015). As is usual in these kinds of reflections on desiderata for theories of truth, the authors begin by pointing out that they do not expect there to be a list of desiderata which gives a final verdict over which theory of truth is the best or the fully correct one. Rather, there is a certain kind of pluralism in the air, suggesting that different authors and different theories may be seen as focussing on different aspects of truth. Horsten and Halbach point out the following desiderata of theories of truth on which they would like to focus. The first desiderata of coherence is the most complex one, which comes in many different forms:

1. A theory of truth should be coherent.

   - *Compositional coherence*: If truth behaves compositionally with respect to conjunction, then it should also do so with respect to disjunction. The reason for this being that there seems to be no reason why compositionality should hold for one but not the other, especially with respect to the possibility of defining them in terms of each other and negation.
   - *Base coherence*: The truth theory should not contradict the base theory. Adding a truth predicate and principles governing it to e.g. an arithmetical theory should not allow us to derive statements, which contradict what can be proven in that arithmetical base theory.
   - *Soundness coherence*: Even though truth theories may be too strong to prove their own soundness, it should be possible to consistently add a soundness claim to the theory. For example, adding the reflection principle $\forall x(Prov(x) \rightarrow Tx)$ to a theory of truth should be possible without leading to inconsistency.
   - *Language coherence*: Adding a truth predicate to the language of the base theory, all principles and axiom schemes should be extended by the truth predicate. In the case of PA this amounts to the fact that the Induction scheme be extended to formulae including the truth predicate.

- *Internal coherence*: The theory should not only be non-trivial but also not internally trivial. This is to say that there should be some formulae $\phi \in \mathcal{L}_T$, s.t. $T\ulcorner\phi\urcorner$ is not derivable.

The next two desiderata which they list are disquotation and compositionality for T. However, these can be subsumed under what we have already discussed above with respect to Leitgeb's list. They do go at greater length to discuss different forms of disquotation (transparency, biconditionals, uniform biconditionals, intersubstitutivity) but we do not need to formulate this in a new desiderata. However, the last two desiderata in their discussion are new with respect to what we have seen so far.

2. Preservation of ordinary reasoning

This desideratum has two dimensions. The first we have already seen above is concerned with the extension of well-known principles to the added truth predicate of the language. This has to do with preserving a kind of ordinary reasoning of the mathematician as they argue that "[o]rdinary reasoning should be taken to include schematic mathematical or syntactic reasoning. So, for instance, in mathematics we are used to subjecting every predicate to mathematical induction. This means that axiomatic truth theories where the truth predicate is not allowed in the induction schema, do not receive a maximal score on this desideratum." (Horsten and Halbach, 2015: p.272).

The other dimension does not have to do with mathematical principles but rather with logical ones. In particular, it is a critique of non-classical approaches motivated by the often-cited remark by Feferman that "[...] nothing like sustained ordinary reasoning can be carried on [...]" (Feferman, 1984: p.95) in non-classical logics (Feferman is in particular talking about partial logics). The requirement here is that the logic underlying the theory of truth should be, at least in principle, applicable informally. Of course we can prove many (meta-)theorems about all kinds of non-classical logics, but it is a different thing to use it proving theorems in some theory informally as is common in mathematics:

> Our ordinary and mathematical reasoning is carried out in classical logic. Reasoning in intuitionistic logic does not come as natural to most of us, but it can be learned without too much difficulty. Reasoning in partial or in paraconsistent logic is a lot less natural still: it is very difficult to learn. Reasoning fluently in even more artificial logic, such as a logic in which certain structural rules are restricted (such as contraction, perhaps), might well be practically impossible. (Horsten and Halbach, 2015: p.272)

This idea of applicability of a logic in formal proofs (or being able to learn to do so) should of course be extended to principles and axioms governing the truth predicate as the authors argue themselves. Too many such axioms, ones which are syntactically too complex or complex with respect to the conditions under which they can be applied imply that the theory does not fulfil this desideratum. Their last desideratum has to do with the philosophical background story of a formal theory of truth:

3. A formal theory of truth ought to capture a philosophical story about which principles are to be restricted and how

Here the idea is that the axioms of the logical system and the axioms for truth should not only be easily learnable for informal reasoning, but they should also be motivated. Such motivation is typically achieved by arguing that the chosen axioms capture or reflect a certain philosophical picture or story. A straightforward example of this are attempts to capture Kripke's fixed point construction from (Kripke, 1975) axiomatically. The philosophical story behind the model-theoretic construction is that it reflects the way a speaker learns about the meaning of the truth-predicate. We start with some ordinary language not containing T. Further, we learn that the truth predicate is applicable to any atomic sentence of the language we have spoken so far. Finally, truth of complex sentences is determined compositionally.

The semantic clauses of Kripke's construction can then be turned into axioms. Thus one may argue that the choice of axioms is motivated by the philosophical story of how a competent speaker learns about the meaning of the truth predicate and how to use it correctly. In general, what the desideratum tells us is that we should have some argument ready in order to defend the choice of axioms and argue for this particular choice over other potential options.

Having gone through and explained various desiderata discussed in the literature, we now want to bring all of them together in a more systematic way. In the following, we offer a way of grouping the desiderata together into three different categories: desiderata regarding the truth predicate, the underlying logic and consistency.

1. **Desiderata regarding T**

    1.1. *Syntactic desiderata*

        1.1.1. T is a predicate of objects referring to sentences, the theory of truth includes a syntax theory providing and governing these objects
        1.1.2. There are no type restrictions applying to T
        1.1.3. Principles and axioms of the base theory are fully extensible by T

    1.2. *Inferential desiderata*

        1.2.1. All instances of the (uniform) T-biconditionals should be derivable
        1.2.2. The choice of axioms for T should reflect some philosophical story or motivation
        1.2.3. Truth should behave compositionally in a way which preserves compositional coherence
        1.2.4. T should be able to fulfil the substitutional roles
        1.2.5. T should be able to fulfil the quantificational roles
        1.2.6. The principles governing T should make it possible to prove the consistency of the base theory

2. **Desiderata regarding the underlying logic**

    2.1. The logic should be classical

    2.2. The outer and inner logic should coincide

    2.3. The logic(s) should be learnable and applicable informally

2.4. If classical logic is restricted, it should not cripple the functional roles of T

3. **Desiderata regarding consistency**

   3.1. The truth theory should be non-trivial

   3.2. The truth theory should be internally non-trivial, i.e. there is some ϕ s.t. $T\ulcorner\phi\urcorner$ is not provable

   3.3. The truth theory should be consistent

   3.4. The truth theory should be ω-consistent

   3.5. The truth theory should not contradict the base theory

   3.6. The truth theory should be consistent with the addition of reflection principles

It is easy to see that some desiderata in the lists are redundant in the sense that they are implied by other desiderata (but not vice versa). For example, a consistent theory of truth is of course also non-trivial. Nevertheless, we prefer to present the desiderata this way. For it may be the case that a theory of truth may not fulfil some desiderata (in order to sustain others). In this case it would then be important to see which of the desiderata implied by the one theory fails to fulfil can still be maintained. So in the example above, if a theory fails to be consistent, we would still like to check whether it is non-trivial and internally non-trivial although these things are consequences of consistency. Further, some desiderata interact with each other across the different groups from 1 to 3. For example, in the presence of a fully classical conditional, the substitutional role of the truth predicate would imply the derivability of all instances of the (uniform) T-biconditionals. Finally, some desiderata are inapplicable to some approaches. If we desire a fully classical theory of truth, the desiderata of not restricting the logic in a way that interferes with the functional roles of T is irrelevant since it is fulfilled in every case.

Especially due to the last phenomenon of some desiderata not being relevant for some approaches, we might wonder what a list of fully relevant and in principle satisfiable list of desiderata looks like. As Tarski's theorem tells us, the main issue with theories of truth is that the functional roles of T cannot be fulfilled non-trivially while working with full classical logic. Assuming consistency, we can only have one or the other: classical logic or the fulfilment of all functional roles of the truth predicate. As a consequence, we device lists of desiderata for each of the disjuncts:

Desiderata for classical approaches

1. **Desiderata regarding T**
   - 1.1. *Syntactic desiderata*
   - 1.2. *Inferential desiderata*
     - 1.2.1. The choice of axioms for T should reflect some philosophical story or motivation
     - 1.2.2. Truth should behave compositionally in a way which preserves compositional coherence
     - 1.2.3. T should be able to fulfil the quantificational role
     - 1.2.4. T should be able to fulfil the substitutional role as well as possible
     - 1.2.5. The theory should prove a wide range of instance of the (uniform) T-biconditionals
     - 1.2.6. The principles governing T should make it possible to prove the consistency of the base theory
2. **Desiderata regarding the underlying logic**
3. **Desiderata regarding consistency**

In the case of syntactic desiderata regarding T and the consistency desiderata, we left out the individual desiderata as they are unchanged compared to the original list above. The main point of this list is that all the desiderata regarding the underlying logic and consistency are enforced, whereas the inferential desiderata regarding T are restricted. In the case of non-classical approaches, we make the opposite move: we focus on the fulfilment of the inferential desiderata and pay the price of giving up some of the items on the lists of logic and consistency.

Desiderata for fully functional approaches

1. **Desiderata regarding T**
2. **Desiderata regarding the underlying logic**
   - 2.1. The outer and inner logic should coincide
   - 2.2. The logic(s) should be learnable and applicable informally
   - 2.3. If classical logic is restricted, it should not cripple the functional roles of T
3. **Desiderata regarding consistency**
   - 3.1. The truth theory should be non-trivial
   - 3.2. The truth theory should not contradict the base theory
   - 3.3. The truth theory should be consistent with the addition of reflection principles

In the desiderata regarding consistency, we can eliminate the requirement that the theory be internally non-trivial since this immediately follows by the non-triviality requirement and the fulfilment of the substitutional role of T. The two lists above collect features which can and should be expected of our best classical or non-classical theories of truth. We now go on to argue that our nontransitive, fully functional approach fares very well with respect to the non-classical list above. In fact, it fares much better than any non-classical alternative on the market.

Beginning with the desiderata regarding T, we can easily see that all desiderata – both syntactic and inferential – are met. Syntactically, we have made no restrictions at all, Induction is extended to the full language and there are no type restrictions. Inferentially, the presence of unrestricted T1 and T2 rules guarantee that the substitutional roles are fulfilled. Picking some suitable classical theory like $\mathsf{UTB}(\mathsf{Z}_2$ as S, our nontransitive theory $\mathsf{NTT}[\mathsf{S}]$ also fulfils the quantificational roles completely. Since $\mathsf{NTT}[\mathsf{UTB}(\mathsf{Z}_2)]$ proves the translation $\Gamma^* \Rightarrow \Delta^*$ of any sequent $\Gamma \Rightarrow \Delta$ provable in Z_2, it emulates second-order quantification as well as possible.

Given the fully classical rules for $\rightarrow$, we can also prove all instances of the uniform T-schema by using T1 and T2. $\mathsf{NTT}[\mathsf{S}]$ includes all compositional principles unrestrictedly and so there are also no compositional coherence issues. Regarding the philosophical story behind the truth theoretic principles, we have chosen a non-classical approach so we do not have to motivate any such restrictions. The motivation with respect to truth is the naive one.

What about the desiderata concerning the underlying logic? Let us start with whether or not the restriction of classical logic interferes in any bad way with the truth principles, s.t. the truth predicate cannot fulfil any of its roles. We argue that there is no such interference. The quantificational part is played mostly be the classical theory embedded in our nontransitive approach, so the restriction of Cut plays no role here. If we 'step out' of the classical theory S and into nontransitive lands, we can still apply the truth theoretic principles unrestrictedly and that is enough to ensure the fulfilment of the substitutional roles. Thus we argue that there is no relevant drawback in the interaction between the truth theoretic principles and the restrictions on classical logic in $\mathsf{NTT}[\mathsf{S}]$.

Horsten and Halbach argued that restrictions on either the logic or the truth theoretic principles should be applicable easily also in informal proofs. They paid special attention to some substructural approaches, naming the noncontractive one, and argued that such approaches would fare worst with respect to this desideratum. The argument here seems to be that we seldomly (or probably never) pay attention to number of times we have used an assumption or whether we have used the multiplicative or additive formulation of an operational rule in an informal mathematical proof. Thus it would be very hard (if not impossible) for the ordinary practitioner to learn how to apply such a logic. To begin with, things are not so bad for the nontransitive case as they are for the noncontractive one. Mathematicians point out a lot of explicit uses of transitivity by constructing lemmata, which are then combined to prove a theorem. The same holds in many cases for uses of modus ponens (in the sense of the elimination rule for $\rightarrow$ discussed in the previous chapter). So this already speaks in favour of the nontransitive approach.

Further, the restriction on Cut rarely comes to play and is easily learnable. It rarely comes into play because the more complex, mathematical proofs of reconstructing proofs of second-order theorems in our theory of truth happen completely within the classical theory S. So we can do business as usual and need not worry about any problems with Cut. It is only when we wish to make unrestricted use of the substitutional role of T that we must not apply Cut. But first, it is not clear what more complex proofs there are in which we need Cut after using instances of T1 and T2 not included in S. Second, even if we need to do so, we can easily check whether the informal proof is sound with respect to NTT[S] by looking at the order of applications of rules.

What about the relation between the inner and the outer logic in our nontransitive approach? Here things are not so straightforward. With respect to classical theories of truth, it is often demanded that the two coincide. Counterexamples to this fact are often given by pointing out that e.g. the theory might derive all instances of excluded middle, yet still proves that some of them are not true. This is also the case with NTT[S]: it proves all instances of $\Rightarrow \phi \vee \neg\phi$, but also $\Rightarrow \neg T(\ulcorner \lambda \vee \neg\lambda \urcorner)$. However, since we deal with an inconsistent (yet non-trivial), non-classical approach, this should come as no surprise and we think that this criterion is unfit to be applied to such non-classical approaches.

However, we still might want to make sense of the idea that (classical) logic should apply across the board - regardless whether we talk about the inner or outer logic. Understood this way, the desideratum is not fulfilled, since NTT[S] is only classical as long as we stay within the boundaries of the classical theory S. Thus classical logic is not applicable in all contexts. Nevertheless, we would like to point out that NTT[S] dissatisfies this desideratum in an independently motivated way and does so very elegantly. The independent motivation comes from understanding that Cut is really a consistency assumption in disguise and that such an assumption is not justified in NT unrestrictedly due to the paradoxes. Our restriction on Cut directly mirrors the requirement of consistency by demanding that both Cut-premises must be derived in a classical and thus consistent theory of truth. Further, the inner and outer logics do coincide over a wide range of cases. Even in NTT (without the embedding of a classical theory of truth), full PAT is preserved so there is no distinction between the logics on the basis of the vocabulary of formulae. Regardless of whether the formula is purely arithmetical or contains the truth predicate – as long as the proof happens within PAT, classical logic can be applied across the board. This feature is further strengthened by embedding classical theories S which then also add instances of T1 and T2, s.t. their conclusions are governed by classical logic as well.

In this context we introduced the distinction between a theory being $\mathcal{L}$-classical or S-classical. It is $\mathcal{L}$-classical iff it proves all sequents of formulae for the full language (in our case $\mathcal{L}_T$) provable in classical logic. S-classicality is obtained if the theory proves all sequents provable in some other theory of truth closed under classical logic. Where $\mathcal{L}$-classicality is well motivated by the idea that logic is language independent, no theory of truth can be fully S-classical with respect to the naive theory of truth NT. Thus also classical theories of truth do not and cannot meet this requirement. Our nontransitive approach is $\mathcal{L}$-classical. There are no restrictions of the rules with respect to language and all classically provable sequents for the full language remain provable. Further, a strong sense of S-classicality can be obtained by embedding a strong classical theory of truth S into our nontransitive approach obtaining $\mathsf{NTT}[\mathsf{S}]$. So where full S-classicality cannot be obtained, our nontransitive approach still preserves a strong, restricted version of this notion.

We argue that having an inner and outer logic, which do not coincide, is a necessity of any nonclassical theory of truth meeting other basic requirements. With a nonclassical approach, we have to have a nonclassical logic at play when it comes to at least some formulae containing the truth predicate, for otherwise the theory would be trivial. However, we also need full classical logic for at least all arithmetical formulae, since we consider preserving classical PA as a necessity. So there needs to be some distinction between logics when it comes to some T-formulae. Our main point here is that $\mathsf{NTT}[\mathsf{S}]$ suffers from this defect as well but (and this is a big but) it does so in a motivated, elegant way, still surpassing all other non-classical approaches known so far.

Finally, we come to the class of desiderata regarding consistency. The infamous Liar derivation using the inversion principles

$$\dfrac{\dfrac{\begin{matrix}\vdots\\ \Rightarrow \lambda, T\ulcorner\lambda\urcorner\end{matrix}}{\Rightarrow T\ulcorner\lambda\urcorner}\,\text{T2} \qquad \dfrac{\begin{matrix}\vdots\\ \lambda, T\ulcorner\lambda\urcorner \Rightarrow\end{matrix}}{T\ulcorner\lambda\urcorner \Rightarrow}\,\text{T1}}{\Rightarrow}\,\text{Cut}$$

shows that if i) fully classical PAT is preserved, ii) T1 and T2 are unrestricted, then $T\ulcorner\lambda\urcorner \Rightarrow$ and $\Rightarrow T\ulcorner\lambda\urcorner$ are provable. Given the classical ¬R-rule (after applying a T-rule), we can then derive $T\ulcorner\lambda\urcorner$ and $\neg T\ulcorner\lambda\urcorner$. So given the desiderata i) and ii), we have to accept some inconsistencies. Given minimal arithmetic, we can then show that the theory is also ω-inconsistent. For this reason, our nontransitive approach should only be judged with respect to the non-triviality criteria and whether the theory contradicts the base theory or not.

We have established early on that $\mathsf{NTT[S]}$ is non-trivial (also internally due to the transparency of the truth predicate) since the paradoxes rely on the assumption that reasoning is transitive after T-inferences are made under which S cannot be closed consistently. Further, we determined that $\mathsf{NTT[S]}$ is $\mathcal{L}_{\mathsf{PA}}$-conservative over S. So if S does not contradict the base-theory, $\mathsf{NTT[S]}$ does not do so either.

Last, $\mathsf{NTT[S]}$ can deal with reflection principles. Consider for example the global reflection principle (see e.g. Kreisel and Lévy, 1968) saying that everything which is provable in our theory is true: $\forall x(Prov(x) \rightarrow Tx)$ where we take Prov to be a standard provability predicate of the theory to which we add the reflection principle. The reflection principle as an axiom tells us that if we can prove something in our original theory, then we may infer that it is true. This is why such reflection principles are often motivated by the conviction that if one endorses a theory, one is committed to the truth of what it proves (see e.g. Feferman, 1991).

The NEC rule allow us to express this trust in our theory as well:

$$\frac{\Rightarrow \phi}{\Rightarrow T\ulcorner\phi\urcorner}\,\text{NEC}$$

Upon a proof of ϕ (with no side formulae in the sequent), we may infer $T\ulcorner\phi\urcorner$. However, NEC and global reflection are incomparable in some sense, since neither of them is derivable in the current setting from the other. NEC is not derivable from global reflection since we can apply the rule to theorems which we have proved using NEC. This is not the case with the reflection principle, since it can only be applied to things provable in the original theory, not generally to what is provable in the theory extended by a reflection principle. On the other hand, NEC does not prove the global reflection principle in its universally quantified form. It does, however, prove all instances of the schema $Prov(\ulcorner\phi\urcorner) \rightarrow T\ulcorner\phi\urcorner$ as can be easily checked by using NEC, T2 and →R.

Every instance of NTT[S] is inconsistent and ω-inconsistent. So the only additional harm a reflection principle could have is to render the theory trivial. However, since the global reflection principle introduces a truth predicate, we should handle it along the lines of T1 and T2. In particular, this means that Cut should be non-applicable after making use of global reflection. Given this modification, it is easy to see that every instance of NTT[S] + global reflection will be non-trivial as well. So we conclude that NTT[S] can deal with reflection principles.

6.2.2 Virtues From Natural Science

So far we have discussed desirable features of formal theories of truth with respect to the functional roles of the truth predicate as well as desiderata one might come up with by reflecting on what a theory of truth should look like. Both of these classes of considerations are specific to the truth predicate and are only discussed within the literature on formal theories of truth. Another set of theoretical virtues sometimes mentioned in the discussion of theories of truth are virtues such as beauty, simplicity, depth etc. which can originally be found in the philosophy of science. There they are used to discuss empirical theories from the natural sciences.

A problem in the literature on formal theories of truth is that these virtues are hardly discussed systematically. Typically, only an open ended list is given together with some vague remarks about which virtues are fulfilled by ones own theory or which of them are not fulfilled by some rival theory; "Scientific theories are compared with respect to how well they fit the evidence, of course, but also with respect to virtues such as strength, simplicity, elegance, and unifying power" (Williamson, 2017: p.334). In contrast to the literature on formal theories of truth, the literature on scientific practice and theory change already has a long history of discussing these matters. So it makes sense to have a look at what has been done within this area of philosophy of science and see what can be transferred to our subject matter. This chapter discusses the most systematic account of such virtues given so far in (Keas, 2018), makes some adjustments, and argues that our nontransitive approach fulfils many of these virtues.

(Williamson, 2017) is the only systematic attempt to discuss a range of theoretical virtues when it comes to picking a classical or non-classical solution to semantic paradoxes (including at least both truth-theoretic and vagueness related paradoxes) we know of. Instead of just discussing Williamson's paper, we choose a different approach and start from a systematic account of theoretical virtues from the literature on virtues in the natural sciences and see what can be used from that for our purposes.

(Keas, 2018) gives a systematic overview over virtues typically discussed in favour or against scientific theories. His virtues are grouped into four groups. Each group contains various virtues, which are ranked or ordered by expansion. So each succedent virtue, in some way, contains and expands or generalises the precedent virtue. We first give Keas' list including some explanation of virtues which might not be self-explanatory, before we argue which virtues are relevant for a formal theory of truth and which are fulfilled by our nontransitive approach.

1. Evidential virtues
   - 1.1. Evidential accuracy: The theory fits the (empirical) evidence well.
   - 1.2. Causal adequacy: The theory includes causal factors, which explain the observed phenomena in question.
   - 1.3. Explanatory depth: The theory is able to answer a wide range of counterfactual questions, i.e. it is able to make correct predictions for a preferably wide range of counterfactual scenarios.
2. Coherential virtues
   - 2.1. Internal consistency: The theory is consistent.
   - 2.2. Internal coherence: The theory is composed into an overall intuitively plausible picture, this excludes e.g. ad-hoc hypotheses
   - 2.3. Universal coherence: The theory fits well with other theories and findings from other research fields (in particular, it does not contradict them)
3. Aesthetic virtues
   - 3.1. Beauty: The theory "[...] evokes aesthetic pleasure in properly functioning and sufficiently informed persons." (Keas, 2018: p.2762)

3.2. Simplicity: Compared to theories explaining the same facts, the theory does so with less theoretical content.

3.3. Unification: Compared to theories with the same amount of theoretical content (or more), the theory explains more facts.

4. Diachronic virtues

4.1. Durability: The theory has survived a history of testing and predicting the correct outcomes of new experiments or observations.

4.2. Fruitfulness: The theory generates new knowledge or insights by unifying other theories or making novel predictions.

4.3. Applicability: The theory allows for applications in the real world, for example by motivating and making possible new technologies.

Before we have a look at individual virtues, we can already see that some virtues play no role for our case of formal theories of truth. *Evidential accuracy* is of course not applicable if we understand it strictly in terms of empirical evidence. For it is hard to see, what empirical observations would be relevant in judging a solution to truth-theoretic paradoxes. One might lift the restriction to specifically empirical data and argue that the phenomena which need to be accounted are the paradoxes. Evidential accuracy then becomes the claim that a formal theory of truth ought to give a solution to all truth-theoretic paradoxes. However, with this understanding, the virtue collapses into a mere consistency or at least non-triviality desideratum we already discussed in the previous sections.

Causal adequacy also does not appear to be particularly relevant simply because there are no causal claims[1] in formal theories of truth. We are dealing with abstract concepts such as truth and abstract objects such as those of arithmetic. Again, one might interpret this similarly to the previous virtue of empirical adequacy as being able to account for all the different phenomena, i.e. the different paradoxes. Again, this would mean that the virtue collapses into a consistency or non-triviality

[1] (Zardini, 2011) seems to invoke some causal role of sentences when explaining his notion of instability. However, this does not seem to be fleshed out and it seems far from clear how to make sense of such causal claims in the domain of self-referential sentences such as the Liar.

desideratum. Another interpretation of the causal adequacy claim is that the theory ought to give a plausible account of how the paradoxes could arise in the first place and how the proposed solution connects to this explanation. This will of course not be a causal relation but still an explanatory one. We find this to be the most plausible and promising interpretation of this virtue. Nevertheless, it then appears that we have already discussed it in the previous section as the desideratum that the theory of truth should capture or be motivated by a certain philosophical picture or idea about truth.

Explanatory depth is the idea that the theory in question should be able to account for a wide range of phenomena – the wider the better. Understood as applied to one particular subject matter such as truth-theoretic paradoxes, this again collapses into the consistency or non-triviality desideratum. However, we might also understand it more liberally in connection with the diachronic virtues of *fruitfulness* and *applicability*. According to this understanding, a theory (or rather the solution to the relevant paradoxes it is built on) should also provide a solution to at least similar paradoxes in other areas. If we have provided a solution to truth-theoretic paradoxes, we might wonder whether an analogous solution can be given for paradoxes of vagueness, of set theory, class theory etc.

In the camp of coherential virtues, we can also relate some of the theoretical virtues to desiderata we have already discussed. Consistency in the sense of the unprovability of a contradiction is a contentious desideratum in assessing formal theories of truth. For some theories are fine with accepting the provability of instances of $\phi \land \neg\phi$. Other theories, dialetheist approaches, even go a step further and embrace these consequences, allowing for true contradictions. Thus demanding of a formal theory of truth to be consistent is too contentious and cannot be judged to be a virtue of formal theories of truth (without further, independent argument).

Internal coherence appears to be strongly related to the desideratum about non-ad-hocness regarding axiom choice discussed in (Horsten and Halbach, 2015). Here we do not see any interesting differences or room for interpretation which would make a separate discussion of this virtue necessary. *Universal coherence* demands that theories should not stand in contradiction with data of other research fields or consequences of their theories. This reminds of the desideratum that a formal theory of truth should not contradict its base theory. The claim is simply widened to theories in different scientific areas as well (empirical or not).

The group of virtues standing out the most when comparing it with the list of desiderata from the previous sections is surely the list of aesthetic virtues. Here we cannot find desiderata to which these virtues could be reduced. Thus they give arguably the most interesting list of virtues and will be discussed below in more detail. When it comes to diachronic virtues, we have already argued that they are in our case of paradoxes reducible to the virtue of *explanatory depth* in our present context of formal theories of truth.

So the only virtues, which cannot be reduced to desiderata for theories of truth we have already discussed are explanatory depth and the set of aesthetic virtues. We now go on to argue that $\mathsf{NTT}[\mathsf{S}]$ fares well with respect to these virtues. We begin with the virtue of explanatory depth. With respect to formal theories of concepts which can lead to paradox, we argued that the best interpretation is that the solution to truth-theoretic paradoxes should be applicable in other domains as well. The nontransitive approach by giving up or restricting the rule of Cut fulfils this virtue very well.

Consider for example paradoxes in naive set theory. Naive set theoretic principles leads to paradoxes such as Russell's paradox by considering the Russell set, which contains all sets that do not contain themselves. Similar to the Liar, we can then deduce that the Russell-set does contain and does not contain itself. By an application of Cut, we can then derive any formula whatsoever. Historically, this has lead to dropping the naive comprehension principle stating that for every property there is a set containing all and exactly those objects satisfying the property. However, (Ripley, 2015a) has shown that by dropping Cut, one can sustain all naive principles for reasoning about sets while remaining non-trivial. So the nontransitive approach is applicable to set-theoretic paradoxes such as Russell's paradox.

Another paradoxical area one might expect a solution to paradoxes to be applicable in, is the domain of vague predicates. Such predicates give rise to the Sorites paradox roughly as follows. Suppose that Tom is not bald as he has clearly a sufficient amount of hair on his head in order to not count as bald. Further, it clearly holds that for any person with n hair being bald, the same person with

$n+1$ hair is bald as well. By instantiating the last principle and applying modus ponens (or some other form of transitive reasoning), we can conclude that Tom is indeed bald. For say that Tom currently has m hair on his head. Then by the general principle above, we can reason that a person with m hair (however large that number may be) is bald as well. Thus Tom is bald. But this contradicts our original judgement and so we land ourselves in a paradox.

(Cobreros et al., 2012) have shown that there is a nontransitive solution to these problems by giving a new semantic framework called strict-tolerant semantics, which can be shown to be sound and complete with respect to a Cut-free formulation of first order logic. So the nontransitive approach is able to deal with paradoxes of vagueness as well. As in the case of the solutions to set-theoretic and truth-theoretic paradoxes, this solution is defended and motivated by arguing that it preserves the originally motivated naive principles, while still blocking the paradoxical conclusion of triviality. Judging from the examples above, we conclude that the nontransitive approach does indeed fulfil the virtue of explanatory depth.

Last, we will discuss whether and if so to what degree, our nontransitive approach satisfies the aesthetic virtues. We begin with the aesthetic virtue of beauty. Keas defines this virtue relative to subjects sufficiently informed about the subject matter. The idea is that if such subjects come across the theory in question or compare it with other theories, they are struck by a certain aesthetic sensation of beauty. This is surely one of the hardest virtues to get a grip on and it is even harder to determine whether a given theory fulfils it or not. In the case of consistency or non-triviality we can give proofs which do not depend on the judgement of an individual. In the case of beauty, nothing like this seems possible.

Nevertheless, we do like to claim that our nontransitive approach is beautiful in certain respects and this beauty is best seen by considering how the approach relates to the overall literature. We think that the beauty of our approach lies in the fact of how little of a restriction has to be done in order to achieve a theory of truth, which is able to have a truth predicate fulfilling both the substitutional and the quantificational functions. The only thing we need to do is to drop Cut whenever we move beyond the boundaries of the classical theory of truth we are currently considering. Its beauty lies in the combination of both strength and simplicity.

This connection to classical theories of truth is another sign of beauty: The nontransitive approach is able to elegantly embed nice, powerful and well understood theories and further improve upon them by adding a transparent (or even substitutional) truth predicate which would not be possible in the classical theory of truth alone. Beauty still lies in the eye of the beholder, so the informed reader will have to judge for himself whether these traits make the nontransitive approach beautiful – but we certainly think it does.

Moving on to the virtue of simplicity, we can argue from similar traits as in the case of beauty. Again, we do not have a precise definition of sufficient and necessary criteria for what it means for a theory to be simple. However, we do believe that by looking at the way the nontransitive theory of truth is constructed, the informed reader will recognize the theory as simple. Theories of truth are given by putting restrictions on some of the principles of our naive theory NT. So whether or not a theory of truth is simple should be judged in terms of whether these restrictions can be regarded as simple. In the case of $\mathsf{NTT}[\mathsf{S}]$, we only need to consider one simple restriction: Blocking Cut whenever its premises are derived using instances of rules not contained in S.

Deciding whether or not the premises of a possible application of Cut are derived within S can sometimes become complex. For example, to construct $\mathsf{UTB}(\mathsf{Z}_2)$, we need to define a translation function between the second-order language of arithmetic $\mathcal{L}_2$ and the first order language of truth $\mathcal{L}_T$. We then need to check whether all applications of T-rules in the derivations of the Cut-premises are applied to formulae in the range of that translation function. Other theories of truth such as CT rely on a range of compositional axioms for all logical constants of the language in order to define what instances of the T-schema their theory admits. Both of these options introduce some complexity compared to the naive theory, which simply includes the T-schema (or all instances of the T-rules) for all formulae of the language and unrestricted compositional principles. As $\mathsf{NTT}[\mathsf{S}]$ is in part defined by the classical theory S, this makes it more complex compared to both a nontransitive approach which simply gives up Cut completely (or at least for formulae containing T) and compared to at least some classical approaches.

However, we do not think that these comparisons of complexity are justified. As

Keas already remarked above in his explanation of simplicity, we need to consider theories, which explain the same facts. We interpret this as saying that we can only compare theories of truth with respect to simplicity or complexity if truth predicate can play the same functional roles in the compared theories of truth. We first need to fix groups of theories which are equally powerful in some relevant respect before considerations of simplicity can become relevant. For one of the easiest theories of truth is to give up the T-rules completely or restrict them to arithmetical formulae. But such a theory is far from being able to fulfil most of the desiderata we expect of an ideal formal theory of truth.

But if we consider only theories of equal or at least similar strength when it comes to the roles of the truth predicate, then $\mathsf{NTT}[\mathsf{S}]$ has no match in the current literature. For there seems to be no other theory of truth, which fulfils both the quantificational role to the degree of proving all translations of provable sequents of Z_2 *and* having a transparent and intersubstitutable truth predicate. Thus there are no theories with which one could compare our nontransitive approach with respect to issues of simplicity or complexity to begin with. Further, we would like to argue that starting from a classical theory of truth, the restriction of Cut is still rather simple and the only restriction we need to impose. We do not require any further or new theoretical terms, the restriction can be formulated purely in terms of notions already available.

Our very last virtue to consider is that of unification. In contrast to simplicity, unification demands that given the same complexity of theoretical content and resources, the theory explains more phenomena. So in our case, we would have to say that a theory with the same complexity of its restrictions fulfils more desiderata of a formal theory of truth. Here we face the problem of finding other theories which we can judge to have the same simplicity as $\mathsf{NTT}[\mathsf{S}]$. Good candidates appear to be other non-classical theories of truth, which attempt to recover classical logic in particular, safe contexts. For example, a typical claim of non-classical approaches is that they can recover full classical logic, whenever we deal with only arithmetical formulae. The non-classical logic is then only applied whenever we consider formulae containing T. Such approaches seem to be comparable in terms of complexity with $\mathsf{NTT}[\mathsf{S}]$ since they too contain some classical subtheory (namely classical arithmetic) and restrict classical logic only when the truth predicate is in play.

The issue however is that these non-classical approaches, which are comparable in terms of complexity, do not even come close to the power of $\mathsf{NTT}[\mathsf{S}]$ in terms of the functional roles the truth predicate can play here. We have noted many times before that non-classical approaches are proof-theoretically much weaker than their classical rivals and these classical approaches are in turn considerably weaker than their embedding into $\mathsf{NTT}[\mathsf{S}]$. So there simply are no theories to compare with $\mathsf{NTT}[\mathsf{S}]$ with respect to unification – at least not in the way understood by Keas demanding that the compared theories are matched in terms of what data they are able to explain. $\mathsf{NTT}[\mathsf{UTB}(\mathsf{Z}_2)]$ is unmatched with respect to the functional roles the truth predicate can play.

We also believe that there is a slightly different, yet related sense of unification, which is one of the most important virtues of our nontransitive approach. Above, we have distinguished between the substitutional and the quantificational function or role of the truth predicate. Historically, i.e. judging from the literature so far, theories of truth are either (more or less) successful in fulfiling one of these roles but not both at same time. Non-classical theories of truth are successful in maintaining a truth predicate which fulfils the substitutional role (or at least are transparent) since the truth-rules are unrestricted. Classical theories of truth are particularly powerful with respect to the quantificational role, being able to relatively interpret stronger subsystems of second-order arithmetic than any non-classical theory so far. Crucially, there is no other theory of truth so far, which *unifies* both of these roles of the truth predicate in one theory. This unification is achieved for the first time by our theory $\mathsf{NTT}[\mathsf{UTB}(\mathsf{Z}_2)]$.

Judging from the considerations and arguments above, we conclude that our nontransitive approach of embedding a classical theory and closing it nontransitively under transparent T-rules is rather successful. It is able to fulfil both the substitutional and the quantificational roles of the truth predicate. It fulfils a wide range of desiderata for theories of truth in particular, which have been discussed in the literature. Last, we have also seen that there is at least some plausible interpretation of theoretical virtues taken from the natural sciences, s.t. $\mathsf{NTT}[\mathsf{S}]$ satisfies these virtues and typically does so more successfully than its rivals.

Chapter 7

Objections And Replies

This section is concerned with objections, which have been put forward in the literature (or could be put forward) against nontransitive theories of truth and how these objections can be met. Some of them have already been published, others have been presented to us by other researchers in some way, yet others we have discovered ourselves. The objections concern various issues in connection with a nontransitive theory (of truth). Some are concerned with the general issue of revising classical logic in general, others see a problem in revising Cut in particular. Others do not focus on the logical revision but on the properties of the theory such as its ω-inconsistency or its claim of preserving classical logic in some important sense. In the following, we have ordered these objections with respect to which aspect of the theory they take to be problematic.

We thereby do not limit ourselves to issues having to do with nontransitive theories of truth, but also consider objections against nontransitive theories of naive validity. In order to discuss these objections, we first introduce such a theory. Of course, the discussion of such objections involves no claim of completeness, the literature on these issues is simply too broad in order to give a comprehensive discussion of such issues. We discuss these objections using the following format: Objections are quoted from representative papers or rephrased from memories of personal correspondence as parts of a conversation and put in *italics*. The replies are then given below.

7.1 Objections Against Revising Classical Logic

7.1.1 Against Revising Classical Logic In General

Objection 1: *"Once we assess logics abductively, it is obvious that classical logic has a head start on its rivals, none of which can match its combination of simplicity and strength. Its strength is particularly clear in propositional logic, since [classical, propositional logic] is Post-complete, in the sense that the only consequence relation properly extending the classical one is trivial (everything follows from anything). First-order classical logic is not Post-complete, but is still significantly stronger than its rivals, at least in the looser scientific sense, as well as being simpler than they are; likewise for natural extensions of it to more expressive languages. In many cases, it is unclear what abductive gains are supposed to compensate us for the loss of strength involved in [restricting] classical logic."* (Williamson, 2017: p.337f)

Reply: We split the objection into two parts. We first discuss the issues arising from the Post-completeness of classical propositional logic and then then issues having to do with classical predicate logic. The notion of strength of a logical theory that Williamson has in mind here has two parts. First, a logical theory S is stronger than some other logical theory S^* iff all theorems of S^* are also theorems of S but not vice versa. He also argues that one should compare the consequence relations of logics in terms of strength, where the consequence relation of S is stronger than that of S^* iff any sequent $\Gamma \Rightarrow \Delta$ provable in S^* is also provable in S but not vice versa. Classical, propositional logic is thus the strongest propositional logic possible, in virtue of its Post-completeness.

The Post-completeness of classical, propositional logic is a fact, so there is no room for discussion here. However, what remains a bit unclear is what exactly follows from this with respect to the choice of a logic. Supposedly, what Williamson has in mind here is to argue along the following lines: Strength is a virtue of logical theories, so when confronted with a weaker or stronger logical theory (salva veritate) one should pick the stronger one. Classical, propositional logic is Post-complete and thus the strongest because it could only be extended into triviality. However, the second premise of this argument is mistaken. Just because a logic cannot be (nontrivially) extended does not make it the strongest one. For it might just be that the logic is incomparable with respect to other logics. In fact, as (Read, 2019) points out, classical propositional logic is not the only Post-complete logic. For example, Abelian logic is Post-complete as well. So it is hard to see how this fact about classical, propositional logic would support its choice.

Let us now consider the issues having to do with classical, first order logic. As Williamson remarks himself, this logic is not Post-complete but nevertheless he argues that it turns out to be the strongest at least in what he calls the looser, scientific sense. By this looser, scientific sense, Williamson means the following. There are cases in which logics are incomparable with respect to the strength criterion. He gives the example that adding propositional quantifiers to classical logic lets one derive a quantified version of excluded middle $\forall p(p \vee \neg p)$. Adding the same to intuitionistic logic lets one derive the negation of that generalisation $\neg\forall p(p \vee \neg p)$. Clearly, none of the two logics is a subtheory of the other (in terms of theorems and in terms of their consequence relations). However, Williamson argues that classical, quantificational propositional logic is to be regarded as stronger compared to the intuitionistic version. This is due to the fact that a generalisation is more informative than a negated universal sentence. For the former tells us that all propositions adhere to some law, whereas the negated version only tells us that there is some counterexample to the universal claim (although intuitionstic logic cannot prove $\neg(p \vee \neg p)$ for any particular p since that would render it inconsistent and thus trivial).

So what are we to make of this looser, scientific sense of strength and Williamson's attribution of this virtue to classical logic? We find it hard to assess this claim with respect to pure logic, which is why we will focus on Williamson's claim that this virtue also holds for classical logic when it is extended to more expressive languages. As before, we will stick to the case of extending classical logic (or classical arithmetic) to a theory of truth. But here we need to make another distinction between extending classical logic (or rather classical arithmetic PA) to PAT by extending the language by a truth predicate or by extending it to NT by extending the language by a truth predicate *and* the set of rules by truth rules (typically T1 and T2).

Consider first the claim that classical logic is better at handling the extension of the language by a truth predicate. Of course, if we just consider the extension of the language, at least most non-classical logicians will be happy to accept classical logic in this case for the theory is perfectly well consistent. It does prove the weak and strong diagonal lemmata but no contradiction from such sentences, since it lacks the necessary truth principles to do so. Things only get ugly when we consider the extension to NT by truth principles. But even in NT, we can consider what happens to PAT in the classical and in the non-classical approach. In the classical approach, PAT is left untouched since the only restrictions which are made are concerned with the truth theoretic principles which are not part of PAT. With most non-classical approaches, however, this is not the case. Their change of classical logic needs to lead into some restriction of PAT. This is because our Liar derivation using the inversion principles tells us that we land in triviality already from the inverted diagonal lemma of PAT together with T1 and T2 and Cut (which are unrestricted on the nonclassical approach). However, our nontransitive approach does not have this problem. We have already seen that all of PAT is preserved since we only need to restrict Cut after making certain T-inferences.

Let us now consider the claim that classical logic is better at handling extensions of principles, such as the truth theoretic principles. Using our distinction between the substitutional and quantificational roles of the truth predicate, we argue as follows. We agree with Williamson with respect to the scientific strength of classical logic when it comes to the quantificational role of the truth predicate. At least so far, investigations into theories of truth have shown classical theories to be significantly stronger, i.e. more successful in capturing the quantificational role of the truth predicate. In this sense we agree with Williamson and his claims of strength for classical logic.

However, things look different when we consider the substitutional role of the truth predicate. Here it is hard to see how a classical theory of truth could possibly fulfil this role. For given classical PAT and a fully intersubstitutable truth predicate (together with assuming transparency after making use of these rules) already commits us to triviality. So we judge that classical logic is less apt for this role of the truth predicate. It is here that we see the strength of non-classical approaches

– or more particularly the strength of our nontransitive approach. So while it is true that classical logic makes for a stronger theory when extended by certain principles, this is not the case for all aspects of such extensions. Our nontransitive approach gives us the best of both worlds while being a non-classical theory of truth. It preserves the quantificational role of the truth predicate made possible in classical theories and adds to it a fully transparent (or even intersubstitutable) truth predicate.

Objection 2: *"The case may indeed be strengthened by reference to the track record of classical logic: it has been tested far more severely than any other logic in the history of science, most notably in the history of mathematics, and has withstood the tests remarkably well." (Williamson, 2017: p.338)*

Reply: The idea behind this objection is clearly to argue that i) a historic track record of successfully testing a theory counts in favour of that theory, and ii) classical logic has the best track record. We do not see any issues with the first claim although there is of course some work to be done to fully flesh out what this testing amounts to, what relevant tests look like. Nevertheless, we can work with some intuitive understanding, looking at established theories which are closed under classical logic (rather than some nonclassical one). So instead of attacking the first claim regarding the relevance of the historic track record of a logic, we focus on the second claim about classical logic having the best track record.

Williamson gives the example of mathematical theories and certainly, classical logic seems to be the most succesful, i.e. most widely applied, logic here. Although there are some experiments of non-classical mathematics (including paraconsistent and even noncontractive mathematics), they form a tiny minority. One might strengthen this argument by pointing out that in mathematical crises such as the search for a safe foundation of mathematics in the face of Russell's paradox has lead to the restriction of set-theoretic principles rather than to a revision of classical logic. Nevertheless, one should be careful with such considerations. One might very well object that it is precisely because non-classical mathematics is such a minority that such arguments turn out to be rather weak. For if we had investigated more non-classical foundations of mathematics, it might have turned out that there are viable alternatives to classical logic and potentially even better ones in certain respects.

But here we do not want to push this point about necessary investigations into non-classical mathematics any further. Rather, we follow a different line of replying to this kind of objection regarding the track record of classical logic, which is similar to the reply to the previous objection. Consider again the distinction between the substitutional and the quantificational role of the truth predicate. We have seen that classical logic is great at letting the truth predicate fulfil its quantificational role but is rather bad at maintaining its substitutional role. Thus the track record of classical logic is not great across all possible extensions or applications of a logic. Sometimes it turns out that a non-classical approach is better at handling certain applications.

One might object that although non-classical logics may be better suited for single purposes, this is not the point of the original objection. The point was not that classical logic is best at every single application or for extending it to any possible theory. Rather, the argument was that comparing all logics, it has the best track record of being applied. There may be some exceptions to such applications, but classical logic is the one with the fewest exceptions. We have no problem with such claims (although they would probably need some more detailed investigation compared to what has been done so far by Williamson and in general), classical logic is probably the single best logic we have. Our point is merely that in order for some tools such as the truth predicate to fulfil their role completely, we need to become a bit non-classical. This is what was achieved by introducing $\mathsf{NTT}[\mathsf{S}]$.

Objection 3: *"Let us return to the choice between restricting classical logic and restricting disquotation. On second thoughts, one might doubt the apparent symmetry between the options. For the constants of classical logic seem to express absolutely fundamental structure. By contrast, the constants at issue in the disquotational principle – the truth predicate, quotation marks – seem to express much less fundamental matters, specific to the phenomenon of language. Thus the comparison between classical logic and disquotation looks analogous to the contrast between a successful theory in fundamental physics and a successful theory in one of the special sciences, such as economics. Suppose that the economic theory is found to be inconsistent with the fundamental physical theory. Faced with the choice as to* *which theory to restrict in order to preserve the other unrestricted, which would you choose? " (Williamson, 2017: p.339)*

Reply: Given the above question, we would give the same answer as Williamson: one should restrict the economic theory rather than the fundamental theory of physics. In our nontransitive approach, we have made the choice of restricting Cut, the principle of transitivity rather than the rules concerning the truth predicate. One might then argue in the same vein as the economics vs. physics example: the transitivity of consequence is more fundamental than the truth rules.

At first glance it looks like the analogy is clear. Just any aspect of the consequence relation of classical logic will be regarded as more fundamental than the truth rules. After all, the truth principles are specific to a particular predicate, whereas the logical principles are independent of any specific vocabulary. One might try to argue that transitivity is of course not given up completely in NTT[S]. Rather only those applications are given up, which are not included in S as a theory already. But then it is still not clear why one would regard these instances of transitivity as less fundamental than the instances of the truth principles extending S. Instead, we argue that the following two considerations undermine the relevance of fundamental vs. less fundamental principles.

First, we have given independent motivation to give up particular instances of Cut. This motivation is twofold. Following Ripley, we argued that understanding that Cut is an extensibility criterion on assertion and denial, we see that it is not clear why it should hold for all extensions of classical logic (or classical arithmetic) by further rules. It might very well be the case that it is not safe for every coherent position to be extended by either the assertion or the denial of ϕ. Second, we argued that Cut is really a rule assuming consistency in disguise. Given a theory such as NT with paradoxes, such an assumption is no longer justified and thus needs to be restricted to a suitable context. Such independent motivations undermine the argument from fundamentality in the absence of similarly strong reasons to abandon the instances of the truth rules missing from a given classical theory of truth.

Second, the fundamentality argument is undermined by considerations about the functional role that the truth predicate can play in a classical or a non-classical theory. When comparing restrictions with respect to which involves more fundamental principles, we ought to compare theories with the same or at least similar strength. Otherwise we would get a clash with the virtue of strength. Consider the case of two physical theories, which differ greatly in strength and explanatory power, such as Newton's theory of gravitation and General Relativity. Suppose we find a contradiction which is relevant to both theories. However, in order to get rid of the contradiction, we would have to restrict more fundamental principles in the case of restricting General Relativity compared to the case of restricting Newton's theory. Surely, we should go for the revision of General Relativity despite the difference in fundamentality of the principles which are to be revised. We take this to be explained by the fact that General Relativity turned out to be much more successful in terms of applications (such as GPS) and explaining empirical data. However, the case looks different if we compare General Relativity and some rival theory which is (at least roughly) equally good in terms of applications and explanations but where the rival theory needs to be restricted in a less fundamental respect. Thus when comparing restrictions in terms of how fundamental the restricted principles are, we should only compare theories which are similar in terms of strength, explanatory power and possibly other virtues.

Our point is now to argue that there is such a difference between classical theories of truth and our nontransitive approach, s.t. a comparison of the restrictions in terms of fundamentality is not apt. As we have argued in replies to other objections and in previous chapters, there is a significant difference with respect to what functional roles the truth predicate can fulfil in a classical or a non-classical theory of truth. Again, where $\mathsf{NTT}[\mathsf{S}]$ is able to let the truth predicate fulfil both its quantificational and substitutional role, this cannot be achieved by a fully classical theory of truth. Given this discrepancy in strength, we argue that the fundamentality criterion should not be applied to compare $\mathsf{NTT}[\mathsf{S}]$ with purely classical theories, since the former is significantly stronger.

Objection 4: *"[Our abductive methodology] recommends us not to compare two logics only in isolation, but also to compare the results of combining each of them with well-confirmed results from outside logic, such as principles of natural science. But now a crucial asymmetry becomes visible. For any complex scientific theory,* *especially one that involves some mathematics, will make heavy use of negation, conjunction, disjunction, the quantifiers, and identity. Thus, restricting classical logic will tend to impose widespread restrictions on its explanatory power, by blocking the derivation of its classical consequences in particular applications.* " (Williamson, 2017: p.340)

Reply: We have not covered the possibility to apply our theory of truth NTT[S] in the natural sciences but this is not special to our theory, since (as far as we know) this has not been done for any theory of truth whatsoever. However, we agree with the general idea of the objection, pointing out that we should compare classical and non-classical logics when applied to specific domains. In our context of formal theories of truth, it seems best to compare them when applied to mathematical theories, such as arithmetic as we have already done from the start.

But comparing classical arithmetic to the arithmetic contained in NTT[S], we see that there are no restrictions whatsoever with respect to the use of negation, conjunction, disjunction etc. We have seen that any proof – including proofs of pure first-order PA and of emulated Z_2 – of any classical theory of truth S can also be given in the exact same way in our nontransitive theory NTT[S]. We only need to invoke a non-classical logic when we make use of instances of truth-rules, which are not contained in S.

But this amounts to no restriction which could become relevant in the comparison with a classical theory as the following dilemma shows. Given some proof t, the classical rival theory S either contains t or it does not. If it contains t, then so does our nontransitive theory NTT[S]. If it does not contain t, then the classical logician has no grounds to claim that the non-classical logician is missing something.

Objection 5: *"That would invalidate a vast array of arguments in mathematics and the rest of science. The dialetheist may respond by permitting instances of disjunctive syllogism in nonparadoxical cases, whichever they are. But, just as before, that still involves a heavy abductive cost across a vast range of ordinary science, by remodeling ordinary scientific explanations in ways that introduce numerous ad hoc elements. The piecemeal reintroduction of instances of missing classical principles involves heavy abductive costs through loss of simplicity and elegance.* " (Williamson, 2017: p.342)

Reply: What Williamson attacks here are typical recovery strategies of non-classical logics to regain some instances of classical rules (and thus theorems). In non-classical theories of truth e.g. it is common to argue that classical logic may be given up, but it still holds for the theory of the base language. So we can use all classical rules on $\phi \in \mathcal{L}_{PA}$ but have to use the non-classical logic on $\phi \in \mathcal{L}_T \setminus \mathcal{L}_{PA}$. This restriction is a syntactic one because it is only concerned with the syntactic form of the premises of a possible inference. At least most (if not all) recovery strategies for classical logic in a non-classical framework work syntactically.

Williamson's critique is now to argue that such an approach has abductive costs, because it is ad hoc in character. Where classical logic simply contains all instances of all classical rules and thus expresses a maximum of generality, the non-classical logician cannot give a general account. Rather, she needs to give up classical logic in order to get around the paradoxes and then reconstruct classical logic for particular cases. Adding instances of classical rules has an ad hoc character, at least when compared to the general account of classical logic.

We have two responses to this kind of objection. First, while it is true that a classical theory of truth is able to maintain the full generality of classical rules, it cannot maintain the truth rules in its full generality. The typical non-classical theory, which Williamson argues against, may not be able to have the classical rules in its full generality but it can still hold on to the truth rules in their full generality. If the non-classical theory needs to reconstruct classical logic, then the classical theory needs to reconstruct the truth rules. So the argument may very well backfire on the proponent of the classical theory, since it involves ad hoc reconstruction as well. One may object by invoking a notion of fundamentality as in an earlier objection, arguing that such reconstruction is less severe in the truth rule case than it is in the logic case. However, we have already argued above that one needs to be careful when invoking such fundamentality notions in order to argue for abductive gains of one theory over another.

Our second response is concerned with our nontransitive theory $\mathsf{NTT}[\mathsf{S}]$ in particular and comes in two parts: First, our restriction of Cut by demanding that both Cut-premises be derive in some classical theory of truth has nothing unsystematic or ad hoc it. Where the typical non-classical logician may need to admit new instances of classical rules when they are needed, we have given a general recipe of obtaining our nontransitive theory NTT[S] as soon as we are presented with a classical theory S. Simply take S, add all missing instances of truth rules and restrict Cut whenever at least one premise was not derived using the original rules in S. Second, one need not interpret our nontransitive theory as a way of recapturing classical logic or a classical theory. Similar to what we remarked in the first response to the overall objection, it is perfectly fine to see the nontransitive approach as a way of recapturing the instances of the truth rules missing from S – rather than recapturing instances of Cut for a fully nontransitive theory of truth. Such an interpretation would render NTT[S] a classical approach to truth and move it outside the scope of Williamson's objection.

We believe that the above reply also serves as a reply to a similar objection with respect to the alleged lack of systematicity of non-classical approaches:

> Consider any other theorem-schema of classical mathematics, formulated with predicate variables like F. Does the non-classical logician simply look at it, and make an educated guess at the minimal mutilation of it to escape counterexamples with vague or semantic predicates in place of the variables? Such an unsystematic, conjectural approach would fall far short of the standards of contemporary mathematics, which provides an enormous accumulating body of theorems ultimately derived from a very small group of first principles [...]." (Williamson, 2018: p.415)

7.1.2 Against Revising Transitivity

Objection 6: *"In general, substructural logics are ill-suited to acting as background logics for science."* (Williamson, 2018: p.413)

Reply: This objection needs some clarification. The starting point of Williamson's dismissive remark is that logic underlying for example a scientific theory should be able to play the role of a closure operator. He identifies three rather uncontroversial features such a closure operator Cn should have (see (Williamson, 2018: p.413)):

1. If $\Gamma \subseteq \Delta$ then $Cn(\Gamma) \subseteq Cn(\Delta)$

2. $\Gamma \subseteq Cn(\Gamma)$

3. $Cn(Cn(\Gamma)) \subseteq Cn(\Gamma)$

The desiderata of a closure operator together entail all typical properties of the classical consequence relation, including transitivity and contraction (assuming that Γ, Δ are sets and not multisets). So the desiderata of the closure operator rules out a substructural (at least nontransitive or noncontractive) consequence relation and so any nontransitive or noncontractive logic. Further, Williamson argues that such a closure operator is important for scientific theories because they are assessed by looking at their consequences (which are elements of the result of applying a closure operator to the theory). But if our logic is substructural, we would have to give up at least one of the desiderata above for a closure operator. Thus, Williamson argues, substructural logics are not apt for backgrounds logics of science.

Our reply to such an objection is reminiscent of the replies above to previous objections: Given some (consistent) classical theory of truth S, we can give a nontransitive theory $\mathsf{NTT}[\mathsf{S}]$ which has the same consequences and the same consequence relation as S as long as we stay within the boundaries of S. It is only when we apply rules, which are not included in S that we need to invoke a different consequence relation. But then we are outside the scope of what any purely classical theory can do. So the comparison of the consequence relations with respect to the above desiderata is not in favour of classical logic. $\mathsf{NTT}[\mathsf{S}]$ and its consequence relation still fulfil the desiderata as long as we work within S. If we go beyond S by applying new instances of the truth rules, the comparison of the consequence relation makes no sense because the classical theory does not and often cannot include these inferences.

Objection 7: *"Imagine someone who says 'I accept that special relativity entails that nothing travels faster than the speed of light, and I accept that special relativity plus the claim that nothing travels faster than the speed of light entail neutrinos do not travel faster than the speed of light, but I do not accept that special relativity entails that neutrinos do not travel faster than the speed of light.' I think that this sort of claim would strike the audience as baffling. Thus, I do not see that denying* *transitivity is any better than denying modus ponens or conditional proof. That is, this structural rule is constitutive of the derivability operator.*" (Scharp, 2013: p.80)

Reply: The argument aims to establish that transitivity is a constitutive feature of the consequence relation (or what Scharp calls the derivability operator). It is not just a desiderata but if we were to drop this feature, the result would be 'baffling' – which we interpret as 'it does not make sense to accept the two consequence claims but reject the conclusion' or 'dropping transitivity makes it hard to even grasp what the speaker is talking about'. We first consider the objection with the kind of example Scharp gives in the quote above and then a possibly more pressing kind of example involving truth theoretic paradoxes.

The quote uses an example regarding positions about the theory of special relativity SR – so a rather well-understood physical theory which we (as typically anyone else) will assume to be consistent. Dropping Cut would then make it possible to assert two claims about SR of the form $\phi \Rightarrow \psi$ and $\psi \Rightarrow \epsilon$ but reject the position $\phi \Rightarrow \epsilon$. This would indeed be baffling since we expect at least physical theories to be closed under Cut and we use such transitive reasoning all the time, especially when researching or explaining such theories. Note first that dropping Cut as a rule does not necessarily entail that one is no longer committed to accept $\phi \Rightarrow \epsilon$. It might very well be that a cut-elimination theorem for SR (or sufficiently strong subtheories thereof) hold s.t. even without Cut there is a proof of $\phi \Rightarrow \epsilon$.

Further (and more importantly), our nontransitive approach would not give up such instances of Cut. Assuming that SR is consistent, there is no reason under our approach to put a restriction on Cut or any other classical feature of the consequence relation. The theory simply counts as an instance in which classical logic was shown to be successfully applicable. So to summarise, our nontransitive approach would make no such restriction on transitivity as in the quoted example necessary.

One may object that our kind of reply only works if the example is taken from a consistent theory. However, the same bafflement and incredulous stare may occur if we deny transitivity for cases in inconsistent theories. For example, it will be hard to understand what a speaker commits himself to if he accepts the position that 'the Liar is true' and that 'If the Liar is true, then $0 = 1$' while not being committed to the position that '$0 = 1$'. Of course, such instances of Cut cannot be maintained even in $\mathsf{NTT[S]}$ on pain of triviality. So such examples still show that a nontransitive account would have to give up on at least some instances of crucial properties of the consequence relation.

However, we would like to argue that the instances of Cut which are missing from $\mathsf{NTT[S]}$ should not raise any bafflement about the commitments of the nontransitivist. This is because the lack of these instances in our nontransitive approach is independently motivated by the proper understanding of Cut. As we have shown in earlier chapters, Cut plays the role of an extensibility constraint on assertion and denial. As such, it commits one to the idea that any coherent position can be coherently extended by either the assertion or the denial of any formula ϕ. In certain contexts such as PA, this commitment does not raise any issues. However, in other contexts such as a theory of naive truth NT, we should expect the extensibility criterion to fail for there are ill-behaved sentences such as the Liar. In such cases we should neither assert nor deny the sentence in question. So since the missing instances of Cut are independently motivated, they should not be reason for any bafflement.

Objection 8: *Blocking paradoxes by rejecting Cut brings with it the rejection of modus ponens as a meta-rule. For if modus ponens would still hold, we could derive any ϕ whatsoever from the premises $\Rightarrow T\ulcorner\kappa\urcorner$ and $\Rightarrow T\ulcorner\kappa\urcorner \to \phi$, which are derivable in the nontransitive approach. Since this inferential move is invalidated, on the typical understanding of logical consequence we are thus asked to accept both premises as true but the conclusion ϕ as false. But this would in effect mean that the conditional has a true antecedent and an untrue consequent, which breaks with the common understanding of the conditional* (see Zardini, 2013: p.579).

Reply: In our objection we follow the response of (Fjellstad, 2016a). In order to understand what is going on in this objection, we need to carefully distinguish between the following two things, which can both be labelled 'modus ponens':

$$\frac{}{\phi, \phi \to \psi \Rightarrow \psi}\,\text{MP} \qquad \frac{\Gamma \Rightarrow \Delta, \phi \qquad \Gamma \Rightarrow \Delta, \phi \to \psi}{\Gamma \Rightarrow \Delta, \psi}\,\text{MMP}$$

Where MP stands for modus ponens and MMP for meta modus ponens. The first is a theorem of a sequent calculus, for example of classical logic which derives all instances of it for a given language. All instances of MP are derivable in NTT, so there is no problem here for the nontransitive approach.

The second one, MMP, is not a theorem but a rule. We have seen that there is a close connection between MMP and Cut. In fact, if the right-hand side of the second Cut-premise is non empty, then MMP makes Cut admissible. Given the Cut premises $\Gamma \Rightarrow \Delta, \phi$ and $\phi, \Gamma \Rightarrow \Delta$, we simply introduce a conditional in the second premise on ϕ and some element $\psi \in \Delta$ and then eliminate the conditional via MMP, giving us the conclusion of an application of Cut. So for example due to the Curry paradox, NTT[S] cannot have all instances of MMP.

What is problematic about the failure of MMP? Following the interpretation of (Fjellstad, 2016a), Zardini has two objections in mind: an inferential and a semantic one. The inferential objection is that there is an incoherence in a logic which accepts MP but not MMP for the two express the same logical idea. To this we reply that the failure of MMP should come as no surprise when investigating a nontransitive logic. For we have already seen that MMP makes Cut admissible given that the right side of the second premise is non empty. So pointing out the failure of MMP is like pointing out the failure of some instances of Cut, which is a trivial observation in a nontransitive logic which is able to deal with the Curry paradox.

The semantic problem Zardini sees in the failure of MMP is that this fact contradicts the meaning of the material conditional. Ordinarily, we would think that if a conditional and its antecedent are true, then so is its consequent. But, so Zardini argues, this cannot be the case in the nontransitive approach since we can prove for a Curry sentence κ the sequents $\Rightarrow T\ulcorner\kappa\urcorner$ and $\Rightarrow T\ulcorner\kappa\urcorner \to \psi$ for any ψ. The reply to this semantic objection is very similar to the inferential one. The truth conditions behind the objection are those of the conditional in classical semantics. But of course classical semantics is transitive, so it has to go. For the truth conditions of the conditional already invoke a form of transitivity which is sufficient to bring back paradox. So the failure of these truth conditions simply amount to the intended failure of transitivity. Further, there is an alternative semantics (typically called strict-tolerant) (see Cobreros et al., 2012) which captures the nontransitive account semantically. So the conditional of the nontransitive approach is well-understood and its non-classicality is motivated since the classical truth conditions for the conditional are transitive.

Objection 9: *Applications of Cut can be shown to correspond to derivations in non-normal form in natural deduction. However, the analysis of Tennant of paradoxes in terms of non-normalizable derivations (so where Cut cannot be eliminated) fails. For we can already construct derivations of absurdity in propositional logic (given additional premises) in a way that the derivation is non-normalisable (involves ineliminable applications of Cut). But surely there are no paradoxes in pure propositional logic* (see Schroeder-Heister and Tranchini, 2017).

Reply: Here we reconstruct the original derivation, which is given in a system of natural deduction, for a sequent calculus system in the straightforward way. In the case of natural deduction, two premises $\phi \to \neg\phi$ and $\neg\phi \to \phi$ are added. In natural deduction, these conditionals can then be used by a standard elimination rule concluding the consequent with the antecedent as a further premise. Since NT does not include an elimination rule for the conditional, we add the initial sequents $\phi \Rightarrow \neg\phi$ and $\neg\phi \Rightarrow \phi$ instead, which can then be used via Cut to infer ϕ from $\neg\phi$ and vice versa. The derivation of the empty sequent (absurdity $\perp$ in natural deduction) then looks as follows:

$$\dfrac{\dfrac{\dfrac{\phi \Rightarrow \phi}{\Rightarrow \phi, \neg\phi}\,\neg\text{R} \qquad \neg\phi \Rightarrow \phi}{\Rightarrow \phi}\,\text{Cut} \qquad \dfrac{\phi \Rightarrow \neg\phi \qquad \dfrac{\phi \Rightarrow \phi}{\neg\phi, \phi \Rightarrow}\,\neg\text{L}}{\phi \Rightarrow}\,\text{Cut}}{\Rightarrow}\,\text{Cut}$$

Of course, this derivation would not be possible in pure propositional logic without the addition of the new initial sequents $\phi \Rightarrow \neg\phi$ and $\neg\phi \Rightarrow \phi$ together with the use of Cut. In natural deduction, this derivation is not normalisable, in sequent calculus the applications of Cut are not eliminable. Schroeder-Heister and Tranchini take this to be a counterexample to the analysis of paradoxicality in terms of non-normalisability (ineliminable applications of Cut) of Tennant, which we discussed in earlier chapters. Tennant in (Tennant, 1982, 1995, 2016) argued that paradoxes are exactly those derivations, which cannot be normalised. Yet the derivation above should not count as a paradox. Schroeder-Heister and Tranchini argue that the derivation above "[...] would thus show that loops are not a feature of the extralogical part, but of the logical part of paradoxical derivations. The looping feature would not depend on the possibility to move, for a certain λ, from λ to $\neg\lambda$ and vice versa, but that we can move, for any formula ϕ, from $\phi \leftrightarrow \neg\phi$ to absurdity.“ (Schroeder-Heister and Tranchini, 2017: p.571, notation slightly changed)

By the logical part of a derivation they simply mean any applications of rules of classical or intuitionistic predicate logic. The extralogical part is then any rule added to the logic in order to reconstruct paradoxes. Instead of adding a powerful background theory able to play the role of a syntax theory, they add paradoxes via piecemeal rules. So for the Liar sentence, one adds a rule which allows to infer λ from $\neg\lambda$ and vice versa. The authors argue that the transition from ϕ to $\neg\phi$ and vice versa is still part of the logical part of the system because it is not concerned with any particular sentence. But then the example above would show that there are paradoxes in what the authors regard as the purely logical part. So the analysis of paradoxes in terms of non-normalisability or ineliminable Cuts fails. This argument has sparked some discussion about which reduction rules (in natural deduction) should be accepted in order to correctly classify paradoxes (see (Tennant, 2016; Schroeder-Heister and Tranchini, 2018)) but we will not be concerned with these discussions here as our response to the supposed problem goes in a different direction.

We argue that there are three things wrong with this objection. First, the reconstruction of paradoxes in terms of extra-logical rules is a toy example but mistaken in giving a general account of paradoxes. When investigating theories of predicates which are prone to paradoxes, one should work with a theory which is strong enough to express some notion of self reference and which thereby generates paradoxical sentences. Working with the theory PAT for example shows that the Liar equivalences $\lambda \leftrightarrow \neg T\ulcorner\lambda\urcorner$ or $l = \ulcorner\neg Tl\urcorner$ are theorems of classical arithmetic. So basic arithmetical principles – closed under classical logic – generate paradoxical sentences. They are not extra-logical.

Second, it is not clear at all why we should regard the sequents $\phi \Rightarrow \neg\phi$ and $\neg\phi \Rightarrow \phi$ (or the respective conditionals) as logical. We do not know of any logical system which includes them and see no reason why they should be added to any logical system. The mere fact that ϕ is a schematic variable does not make it a logically valid sequent or conditional.

Third, we have already seen that there are independent and better reasons to drop Tennant's analysis of paradoxes according to which any derivation with an ineliminable application of Cut is a paradox. We only need to extend our system of classical logic by arithmetical principles such as Induction. This renders Cut-elimination for finite derivations impossible and so we get ineliminable applications of Cut. However, noone would regard these derivations involving Induction as paradoxes. So the analysis is mistaken anyway but we should not believe so on the basis of the above derivation.

7.2 Objections Against The Treatment Of Validity

Before we consider objections against the nontransitive response to what is now known as the validity-Curry (vCurry) paradox, we need to show what this paradox is all about. (Beall and Murzi, 2013) introduced a paradox similar to the original Curry paradox, but instead of being formulated via the unary truth predicate, it makes use of a two-place validity predicate. We extend our base language $\mathcal{L}_{PA}$ by the two-place predicate $Val(x, y)$ s.t. $Val(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner)$ is a formula iff ϕ and ψ are. The rest of the definitions remain the same. The result of extending our base language that way is called our language of validity $\mathcal{L}_V$. PAV is then the theory of PA formulated in $\mathcal{L}_V$. NV, our theory of naive validity, is given by all rules of PAV plus the following two rules VP and VD for Val:

$$\frac{\phi \Rightarrow \psi}{\Rightarrow Val(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner)}\text{VP} \qquad \frac{}{\phi, Val(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner) \Rightarrow \psi}\text{VD}$$

We can then obtain the vCurry sentence $\nu \leftrightarrow Val(\ulcorner\nu\urcorner, \ulcorner 0 = 1\urcorner)$ by the weak diagonal lemma just as in the case of the truth theoretic paradoxes. The rules for the validity predicate are not invertible (simply because VD is a zero-premise rule), so we cannot give the structurally identical derivation as in the original Curry paradox.

However, the conditional is still invertible, so we know that $\nu \Rightarrow Val(\ulcorner\nu\urcorner, \ulcorner 0 = 1\urcorner)$ and $Val(\ulcorner\nu\urcorner, \ulcorner 0 = 1\urcorner) \Rightarrow \nu$ are derivable. We can then derive $\Rightarrow 0 = 1$ as follows:

$$\dfrac{\dfrac{\dfrac{\nu \Rightarrow Val(\ulcorner\nu\urcorner,\ulcorner 0=1\urcorner) \qquad \dfrac{}{\nu, Val(\ulcorner\nu\urcorner,\ulcorner 0=1\urcorner) \Rightarrow 0=1}\text{VD}}{\nu \Rightarrow 0=1}\text{Cut}}{\Rightarrow Val(\ulcorner\nu\urcorner,\ulcorner 0=1\urcorner)}\text{VP} \qquad \begin{matrix}\vdots \\ Val(\ulcorner\nu\urcorner,\ulcorner 0=1\urcorner) \Rightarrow \nu\end{matrix}}{\Rightarrow \nu}\text{Cut}$$

$$\dfrac{\begin{matrix}\vdots \\ \Rightarrow \nu\end{matrix} \qquad \begin{matrix}\vdots \\ \nu \Rightarrow 0=1\end{matrix}}{\Rightarrow 0=1}\text{Cut}$$

Where the original Curry paradox involved use of the rules for the conditional (at least in the PAT part of the derivation), no such rules need to be used in the case of vCurry. This fact is often taken to motivate a substructural (i.e. noncontractive or nontransitive) approach to paradoxes over other non-classical approaches. For one would expect the Liar, Curry and vCurry to have the same solution. However, structural non-classical approaches work by restricting the operational rules. Since no such rules are present in the derivation of triviality via vCurry, this is not an option.

Given the argument just stated, vCurry leads to favour substructural over structural non-classical approaches but it does little to favour them over classical approaches. Just as full classical logic may be retained by restricting the truth rules, the same could in principle be done with the validity rules. But here we will not be concerned with this possibility. In the following, we discuss objections against both the supposed motivation for substructural approaches and the nontransitive solution to vCurry.

Objection 10: *Let ν be a vCurry-sentence of your choice. Then in the nontransitive approach we can still derive $\nu \Rightarrow \phi$. Using one of the validity-rules, we can conclude that $Val(\ulcorner\nu\urcorner, \ulcorner\phi\urcorner)$ is true. Truth should be closed under logical equivalence. If so, we can conclude that ν is true as well. However, the fact that absurdity (or any formula of the language whatsoever) follows from a true sentence is what makes a paradox. So the nontransitive approach has not discharged the paradox at all.* (see Shapiro, 2013: p.104)

Reply: We take it that Shapiro's argument roughly has the following structure: i) The problem with paradoxes is that they establish that something false (including the possibility of any formula of the language) follows from something true. ii) The nontransitive approach still proves that anything follows from a true sentence such as the Curry sentence. Therefore, the problem of paradoxes is not solved by the nontransitive approach.

Let us consider the second premise of the argument first. What Shapiro has in mind is that both $\Rightarrow \nu$ and $\nu \Rightarrow \psi$ are provable (where ψ again is arbitrary). We agree that provability should entail truth (speaking now on the meta-level). Thus since ν is provable, it is true. But the second sequent tells us that any arbitrary ψ follows from ν. So any arbitrary formula ψ, and thus something false, follows from ν, which is something true.

When it comes to the first premise, it seems that Shapiro seems to invoke at least a similar definition to what we used as a starting point for our investigation into paradoxes in an earlier chapter. There the idea was, roughly, that a paradox is an argument which leads one from (supposedly) true premises and valid reasoning to a false or unacceptable conclusion. Shapiro's claim then is that what remains derivable in the nontransitive approach about the Curry sentence, namely $\Rightarrow \nu$ and $\nu \Rightarrow \psi$, still fit this definition of a paradox and so the paradox is not solved at all.

However, we want to argue that what is derivable about the Curry sentence, or any other formula of the language, does not fit this intuitive picture of what makes a paradox. Shapiro's mistake, it appears to us, is that he takes the truth of the premises to be sufficient to constitute a paradox while not paying attention to the fact that the reasoning needs to be valid as well. It is this point which marks the crucial difference between the original Curry paradox and what remains provable in $\mathsf{NTT}[\mathsf{S}]$. Whereas in the original we supposed that the application of Cut concluding $\Rightarrow \psi$ was valid, we know better in the nontransitive approach. The move from 'If ϕ is true, then it entails ψ' and 'ϕ is true' to 'ψ is true' may be valid in classical logic due to the leading principle of truth preservation – but this of course ceases to be the case in certain non-classical logics, including our nontransitive one. So the nontransitive approach does indeed block the paradoxes, just not on the level of the truth of premises but on the level of the validity of the involved reasoning.

Objection 11: *There is no paradox of logical validity. The construction of the v-Curry sentence relies on some machinery of expressing self-reference, typically Gödel coding and the diagonal lemma. But such resources cannot be regarded as logical. In the case of* PA *or similar theories, they are arithmetical. If the Validity rules are supposed to capture logical validity, they cannot be applied to the v-Curry sentence, since a derivation of the latter involves not only logical but also arithmetical resources. So there is no paradox of logical validity. Of course, this is no objection against the nontransitive treatment of the paradox. Rather, it weakens the nontransitive approach more generally, by showing that one of its motivations – being able to solve both truth-theoretic and validity paradoxes – is void.* (see Cook, 2014)

Reply: Recall the derivation from above of the empty sequent from the v-Curry sentence:

$$\dfrac{\dfrac{\dfrac{\nu \Rightarrow Val(\ulcorner \nu \urcorner, \ulcorner 0=1 \urcorner) \qquad \dfrac{}{\nu, Val(\ulcorner \nu \urcorner, \ulcorner 0=1 \urcorner) \Rightarrow 0=1}\,\text{VD}}{\nu \Rightarrow 0=1}\,\text{Cut}}{\Rightarrow Val(\ulcorner \nu \urcorner, \ulcorner 0=1 \urcorner)}\,\text{VP} \qquad \begin{array}{c}\vdots \\ Val(\ulcorner \nu \urcorner, \ulcorner 0=1 \urcorner) \Rightarrow \nu\end{array}}{\Rightarrow \nu}\,\text{Cut}$$

$$\dfrac{\begin{array}{c}\vdots \\ \Rightarrow \nu\end{array} \qquad \begin{array}{c}\vdots \\ \nu \Rightarrow 0=1\end{array}}{\Rightarrow 0=1}\,\text{Cut}$$

In this derivation we need to apply the validity rules VP and VD to the vCurry sentence. However, if the validity predicate and the rules governing it are supposed to capture logical validity, then no such application should be possible. For the proof of the sequent $\kappa \Rightarrow \psi$ for example involves not only logical but also arithmetical rules which are necessary to derive the instance of the weak diagonal lemma that is the v-Curry sentence. But if the validity rules are not applicable to anything we derive with the help of arithmetical rules but only with logical rules, then there is no paradox. This is because pure logic is too weak to generate any kind of paradoxes.

Of course, there being no paradox is not in itself a problem for an approach to such paradoxes. We take it that the best objection in this vein is to argue that it undermines the motivation for nontransitive or noncontractive logics. (Beall and

Murzi, 2013) argue that the paradox makes necessary a noncontractive approach, (Ripley, 2013a) argues in favour of his nontransitive approach by pointing out that dropping Cut blocks the derivation of the empty sequent from vCurry. If there is indeed a validity paradox, then this speaks in favour of substructural approaches over other non-classical but fully structural approaches. This is because it is often seen as a virtue for a solution to the paradoxes to be applicable in other domains as well. However, in order to arrive at triviality from the vCurry sentence one only needs to invoke the validity rules and structural rules. Both would still be present in structural non-classical approaches such as K_3 or LP which restrict the rules of the conditional and negation. So if it turns out that there is no threat from v-Curry due to a misunderstanding of the validity rules, this would take away some considerate support in favour of substructural approaches to paradox.

We agree with Cook's analysis that the paradox, presented as a paradox of logical validity, is mistaken. If the validity rules are supposed to capture logical validity, then they cannot be applicable to paradoxical sentences which are generated by nonlogical rules such as arithmetical ones. However, we do not believe that this undermines the motivation in favour of substructural approaches completely. Although correct in his analysis, Cook misses the more general point about paradoxes which are similar to the supposed v-Curry. What is crucial in the original argument is not that we consider logical validity. Rather, what makes the substructural approach necessary is that we can circumvent the use of logical constants such as $\rightarrow$ or $\neg$ by the introduction of certain two-place predicates. Together with some plausible rules for this predicate, there are derivations of the empty sequent which only consist of structural and VP-/VD- inferences but make no use of rules for logical constants.

To strengthen our point, we show that the validity rules VP and VD are equivalent to the classical rules for the conditional if we translate $Val(\ulcorner\phi\urcorner,\ulcorner\psi\urcorner)$ as $\phi \rightarrow \psi$ and vice versa given Cut. Consider first the rule of VP:

$$\frac{\phi \Rightarrow \psi}{\Rightarrow Val(\ulcorner\phi\urcorner,\ulcorner\psi\urcorner)}\ \text{VP}$$

Here it is easy to see that this corresponds directly to the right sequent rule for the conditional $\rightarrow$R (under the restriction that there are no side formulae). The same thing can be shown for VD although it is not as obvious:

$$\frac{}{\phi, Val(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner) \Rightarrow \psi}\ \text{VD}$$

This rule corresponds to the left conditional rule →L given Cut which can be shown as follows. Rewriting VD with the conditional instead of the validity predicate gives us modus ponens as a theorem

$$\frac{}{\phi, \phi \to \psi \Rightarrow \psi}$$

Suppose that we are given the premises $\Gamma \Rightarrow \Delta, \phi$ and $\psi, \Gamma \Rightarrow \Delta$. We can then establish the conclusion of an application of →L as follows:

$$\dfrac{\Gamma \Rightarrow \Delta, \phi \qquad \dfrac{\phi, \phi \to \psi \Rightarrow \psi \qquad \psi, \Gamma \Rightarrow \Delta}{\phi, \phi \to \psi, \Gamma \Rightarrow \Delta}\ \text{Cut}}{\phi \to \psi, \Gamma \Rightarrow \Delta}\ \text{Cut}$$

So what is really going on in the vCurry paradox is that the role of the conditional in the usual Curry paradox is played by the validity predicate. But in order to play this role, a given predicate need not represent validity, let alone logical validity. It only needs to be governed by rules which are strong enough to mimick the rules for the material conditional. In the following we explore some alternative predicates and rules for them. By this we show that although vCurry where validity is understood as logical validity might fail, there are still other interesting paradoxes which show the same structural behaviour of only involving VP, VD and structural rules.

A first idea might be to introduce a predicate Imp for implication expressing the conditional as a predicate. Then $Imp(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner)$ is understood as 'ϕ implies ψ'. Using diagonalisation and the straightforward rules for Imp following the scheme of the rules for Val, we could then give a new version of the vCurry paradox. However, we doubt that such a paradox would have the wanted effect. For if the predicate is simply supposed to express the material conditional, then the proponent of a nonclassical logic which restricts the classical rules for → has every right and motivation to restrict the rules for Imp as well. So this paradox would raise no issue for the structural non-classical logician or support the case of the substructural approach.

Although the strategy might fail for Imp, we believe that there are other predicates for which our strategy is successful. When working in a formal calculus, validity and truth are typically notions of our metalanguage. So it makes sense to express other predicates in the object language which are typically only available in the metalanguage. Consider the predicate $Ent(\ulcorner \bigwedge \Gamma \urcorner, \ulcorner \bigvee \Delta \urcorner)$, which is to be read as 'the conjunction of formulae in Γ entails the disjunction of formulae in Δ'. In other words, Ent represents the intended reading of a sequent $\Gamma \Rightarrow \Delta$ in the object language of the formulae in sequents. We can then give the following rules EP and ED for Ent, which are analogous to the rules for Val in the original vCurry:

$$\frac{}{\Gamma, Ent(\ulcorner \bigwedge \Gamma \urcorner, \ulcorner \bigvee \Delta \urcorner) \Rightarrow \Delta}\text{ ED} \qquad \frac{\Gamma \Rightarrow \Delta}{\Rightarrow Ent(\ulcorner \bigwedge \Gamma \urcorner, \ulcorner \bigvee \Delta \urcorner)}\text{ EP}$$

Consider then the entailment Curry sentence eCurry: $\epsilon \leftrightarrow Ent(\ulcorner \epsilon \urcorner, \ulcorner 0 = 1 \urcorner)$. The rules for Ent are not invertible, so we will have to use some additional Cuts compared to the short derivation of the empty sequent via the common Curry sentence. Nevertheless, the rules for the conditional in the instance of the weak diagonal lemma are invertible. So we know that $\epsilon \Rightarrow Ent(\ulcorner \epsilon \urcorner, \ulcorner 0 = 1 \urcorner)$ and $Ent(\ulcorner \epsilon \urcorner, \ulcorner 0 = 1 \urcorner) \Rightarrow \epsilon$ are provable. We can then prove $\Rightarrow 0 = 1$ as follows:

$$\dfrac{\dfrac{\dfrac{\begin{matrix}\vdots\\ \epsilon \Rightarrow Ent(\ulcorner \epsilon \urcorner, \ulcorner 0 = 1 \urcorner)\end{matrix} \qquad \dfrac{}{\epsilon, Ent(\ulcorner \epsilon \urcorner, \ulcorner 0 = 1 \urcorner) \Rightarrow 0 = 1}\text{ ED}}{\epsilon \Rightarrow 0 = 1}\text{ Cut}}{\Rightarrow Ent(\ulcorner \epsilon \urcorner, \ulcorner 0 = 1 \urcorner)}\text{ EP} \qquad \begin{matrix}\vdots\\ Ent(\ulcorner \epsilon \urcorner, \ulcorner 0 = 1 \urcorner) \Rightarrow \epsilon\end{matrix}}{\Rightarrow \epsilon}\text{ Cut}$$

$$\dfrac{\begin{matrix}\vdots\\ \Rightarrow \epsilon\end{matrix} \qquad \begin{matrix}\vdots\\ \epsilon \Rightarrow 0 = 1\end{matrix}}{\Rightarrow 0 = 1}\text{ Cut}$$

We argue that eCurry, in contrast to vCurry, is able to motivate a substructural logic and overcomes Cook's objection. Let's start with the motivation. In the derivation of a falsity such as $0 = 1$, we only made use of diagonalisation with the Ent predicate, ED, EP and the structural rule of Cut (applications of Contraction are implicit in the context-sharing formulation of Cut and the use of sets instead of multisets). However, there was no use of any rules governing logical constants, so a structural non-classical approach is no option. The only way to keep the naive rules for the entailment predicate is to go substructural.

Does Cook's objection against vCurry have an analogue for eCurry? We think not. With vCurry the problem was that we cannot apply VP and VD to the vCurry sentence because it is obtained by extralogical resources. So if VP and VD are understood as governing logical validity, they cannot be applied to the vCurry sentence. But no such problem arises in the case of eCurry. EP and ED are not limited to formulae derived using only logical resources. The intended interpretation of Ent covers all formulae in any sequent derivable in a given sequent calculus. So via eCurry we get a motivation for substructural approaches compared to other non-classical options while getting around Cook's objection.

To be fair to Cook, he does anticipate a version of our reply and concludes himself that although there may be no paradox of *logical* validity, other interpretations of Val as analytic or metaphysical necessity will do the trick: "Thus, there are paradoxes that can be formulated in terms of important understandings of validity. But there are no paradoxes that plague the notion of logical validity." (Cook, 2014: p.466). We tried to make a similar point, introducing a predicate for entailment rather than interpreting Val differently. The lesson remains the same: There are paradoxes which require substructural logics if the naive rules for certain predicates are to be kept.

7.3 Objections Against The Recapturing Strategy

Objection 12: *You spend a great amount of resources in order to strengthen a nonclassical theory of truth with respect to its proof-theoretic power, such that it becomes non-conservative over the base theory. But especially nonclassical theorists about truth tend to be deflationists and many deflationists see themselves committed to being conservative over the base theory. So it makes more sense for the nontransitive theorist to simply admit Cuts on arithmetical formulae, getting around the* *issue of Cut in* PA *while remaining proof-theoretically conservative.*[1]

Reply: We reconstruct the argument as follows: i) Non-classical logicians (including substructural ones) tend to be deflationists, ii) If one is a deflationist, one is committed to a theory of truth conservative over the base theory, iii) NTT[S], given some suitably strong theory S, is a non-classical theory of truth over the base theory PA. i) - iii) then get us into trouble because NTT[S] is non-conservative although it should be due to i) and ii). Denying iii) is no option since it merely reflects a straightforward theorem about our nontransitive approach. Rather, the problem with the argument lies in premise ii).

There is a long tradition of interpreting deflationism about truth and its claims about the metaphysical lightness or insubstantiality of truth (see e.g. Horwich, 1998) to be committed to a conservative theory of truth. The idea here is typically that if the truth predicate plays no substantial role but is only a tool, then it should not let us prove new facts about the base theory. However, we already argued in the earlier chapter on the functional roles of truth that this understanding of the truth predicate is mistaken. It is precisely because of the functional roles of the truth predicate that we should expect our theory of truth to be non-conservative.

Arguments for non-conservativity from deflationism typically focus on the substitutional role of the truth predicate (see e.g. Cieslinski, 2017). Great emphasis is put on the desideratum that the truth predicate ought to be at least transparent and possibly even fully intersubstitutable. What is neglected is the quantificational role of the truth predicate, which allows one to mimick second-order quantification. Given the technical results of earlier chapters, this functional role of truth cannot be denied. So if deflationists want to embrace the truth predicate due to the roles it plays, they should also embrace the quantificational role. But if so, one should not want or expect ones theory of truth to be conservative. In contrast, one should rather expect it to be non-conservative. For just as we do not expect the result of adding second-order quantification and suitable rules to a theory to be conservative over the original theory, we should not expect the result of adding suitable truth rules to the theory to be conservative either:

---

[1] I thank Lucas Rosenblatt for pressing this objection in private correspondence.

> [...] if a truth predicate can serve as a device for mimicking higher-order quantification, as we submit, then it's neither bad nor puzzling that the truth systems that are up to the task are not conservative over their respective base theories. It is to be expected. (Picollo and Schindler, 2019: p.347)

Objection 13: *Typically, when one tries to recover instances of classical rules in a non-classical logic, one does so for the language of the base theory. This fits the idea that it is safe to apply the classical rules as long as no semantic vocabulary is in play. To achieve the same for the nontransitive approach, one simply needs to add Cut for arithmetical formulae. Thus the whole idea of closing classical theories of truth nontransitively under remaining instances of truth-rules appears to be unnecessary for the business of recapturing classical logic.*[2]

Reply: There are two things we would like to say as a reply to this objection. First, we suspect that the recovery of classical inferences only for formulae of the base language is a consequence of the limitations of such approaches in the literature and not an intention. We take it that if other non-classical approaches had straightforward ways to recover more classical inferences, then they would do so. After all, the restriction of classical logic is typically not independent of the matters of paradox[3] but is seen as a price which one has to pay in order to keep the naive semantic principles such as the truth rules. If so, recapturing as much of classical logic as possible (at least while maintaining some other desiderata) seems desirable for other non-classical approaches as well. It is not that they only want to keep classical logic for arithmetical formulae, they simply lack a systematic way of recapturing more of classical logic.

But even if many non-classical logicians were to put their foot down and claim that classical logic is only to be recovered for the base language, there are independent reasons as why one should not live with this limitation. Classical logic

[2] I thank Lucas Rosenblatt for pressing this objection in private correspondence.

[3] It might be for some authors such as Graham Priest, who does not seem to take classical logic as a starting point but adopts paraconsistent logics for different reasons but sees the logic confirmed due to the fact that it is able to deal with the paradoxes. However, since such approaches are not concerned with recapturing classical logic, they are not relevant here.

is extremely powerful and especially successful in letting the truth predicate fulfil its quantificational role. So in the current absence of a non-classical alternative, we should recover at least as much of classical logic as is necessary in order for the truth predicate to fulfil that quantificational role. Our nontransitive approach fulfils this desideratum, other nonclassical approaches which only recover classical logic for arithmetical formulae do not.

The second response we want to give changes the perspective a bit. As was remarked before, $\mathsf{NTT}[\mathsf{S}]$ need not be understood as a non-classical approach, which tries to recover many classical inferences by embedding a classical theory of truth. We can also understand it as a classical approach, which tries to recover those instances of the truth rules, which cannot, on pain of triviality, be included in the original classical theory. Understanding our nontransitive approach as a way of recapturing the missing instances of T1 and T2 for classical theories puts it outside the scope of the original objection claiming that classical logic should only be recovered for the arithmetical language.

Chapter 8

Conclusion

In this chapter we summarise the dissertation by giving an overview over its advancements in applying nontransitive logics to formal theories of truth compared to the research which has been done so far.

1. We formulated a nontransitive theory of truth over the arithmetical base theory PAT, which is strong enough to prove the existence of self-referential phenomena both in the form of sentences (weak diagonalisation) and terms (strong diagonalisation). This is sufficient to express all paradoxes typically discussed in the literature without adding them in a piecemeal way by additional rules.

2. We showed how to reason with paradoxical sentences generated from an arithmetical theory in a sequent calculus for naive truth NT.

3. The nontransitive theory of truth NTT is formulated with unrestricted compositional principles as rules, which are strong enough to prove the compositional axioms for truth as universally quantified sentences.

4. We showed that a restricted rule of Cut is necessary especially in the presence of powerful tools such as Induction or the truth predicate.

5. We discussed motivations and reasons from the literature to give up or restrict Cut and argued that some of them are inadequate with respect to mathematical theories such as PA or ZF.

6. A new motivation to restrict Cut was found by showing that Cut is equivalent to a rule expressing a consistency assumption. Such an assumption, however, is not justified when reasoning in inconsistent theories. Thus Cut should only be applicable if the premises have been derived in a consistent theory.

7. Based on this motivation to restrict Cut, various ideas of axiomatising this restriction were discussed. Typical axiomatisations of a consisency condition via default logics by introducing defaults (consistency checks) are not viable for the case of formal theories of truth over arithmetic due to issues of complexity and triviality. As a simple alternative, we introduced a restricted Cut rule, which is only applicable if both of its premises are derived using only instances of rules of some consistent theory of truth S.

8. The restricted Cut rule builds on previous research especially on classical theories of truth by embedding such a theory S into a nontransitive setting while preserving it completely, leading to a theory $\mathsf{NTT}[\mathsf{S}]$. Strong instances such as $\mathsf{NTT}[\mathsf{UTB}(\mathsf{Z}_2)]$ can be shown to fulfil both the quantificational and the substitutional role of the truth predicate. It proves all translated sequents of Z_2, proves $T\ulcorner\phi\urcorner$ iff it proves ϕ and proves all instances of the (uniform) T-schema.

9. We showed that the same restriction we used for Cut cannot be applied to any other rule of NT alone in order to obtain a nontrivial theory of truth. Given a classical theory of truth S, Cut is the only rule which needs to be applied after an application of an instances of T1 or T2 not contained in S in all paradoxes.

10. We provided a systematic analysis of desiderata for theories of truth in the literature and argued that $\mathsf{NTT}[\mathsf{S}]$ fulfils the set suitable for fully functional, non-classical approaches very well. Further, we gave an overview of theoretical virtues from the natural sciences and argued that many but not all of these virtues are reducible to the desiderata. Lastly, we argued that $\mathsf{NTT}[\mathsf{S}]$ fulfils these virtues as well.

11. Finally, we surveyed and responded to a number of objections against the

revision of classical logic in general, the restriction of transitivity in particular as well as against the nontransitive treatment of pradoxes of validity by Williamson, Shapiro, Zardini, Cook etc.

Despite the advancements that were made in this dissertation, there are of course open questions and problems, which should be considered in the future:

1. How does a model theory look for $\mathsf{NTT}[\mathsf{S}]$? We have not investigated the model theoretic side of things at all due to the focus on the proof theoretic representation of paradoxes and transitivity. Here we would like to note again that (Cobreros et al., 2012) introduced a model-theoretic approach for nontransitive theories. They showed soundness and completeness for a nontransitive sequent calculus for truth and three valued Kripke-models. Due to the embedding of a classical theory S, a straightforward idea would be to create an inner and an outer model: The inner model is simply a classical model for S, the outer model is a three valued Kripke-model for sequents derived using rules not contained in S.

2. In Chapter 5, we discussed the possibility of obtaining a nontrivial theory of truth by restricting the rules of Contraction in a similar way as in $\mathsf{NTT}[\mathsf{S}]$. Contraction is only applicable, if its premise was derived using only rules contained in S. We showed that this cannot work as long as the multiplicative version of $\to$R is present but conjectured that it might be possible given the additive version of $\to$R. This conjecture needs to be checked, as well as the possibly resulting noncontractive theory of truth.

3. We argued that $\mathsf{NTT}[\mathsf{S}]$ is a better theory of truth than both classical and non-classical alternatives on the market. We expect that one of the biggest objections, especially from the classical camp, will be that $\mathsf{NTT}[\mathsf{S}]$ is inconsistent in the sense of proving $\Rightarrow \phi$ and $\Rightarrow \neg\phi$ for some $\phi \in \mathcal{L}$. It would be nice to defend $\mathsf{NTT}[\mathsf{S}]$ by giving a detailed and plausible interpretation of these inconsistencies. We also think that it makes sense to flesh out an argument as to which success with respect to the functional desiderata proof theoretic strength and proving all instances of the uniform T-schema overwrites the

need for fulfilling philosophical goals like being consistent. But this needs to be fleshed out in more detail.

Bibliography

Aldo Antonelli. *Grounded Consequence for Defeasible Logic*. Cambridge University Press, 2005.

Eduardo Alejandro Barrio, Federico Pailos, and Damian Szmuc. A recovery operator for nontransitive approaches. *Review of Symbolic Logic*, 2018.

Jc Beall and Julien Murzi. Two flavors of Curry's paradox. *Journal of Philosophy*, 110(3):143–165, 2013.

Katalin Bimbó. *Proof Theory: Sequent Calculi and Related Formalisms*. Chapman and Hall/CRC, 2014.

Piero A Bonatti and Nicola Olivetti. A sequent calculus for skeptical default logic. In *International Conference on Automated Reasoning with Analytic Tableaux and Related Methods*, pages 107–121. Springer, 1997.

George S Boolos, John P Burgess, and Richard C Jeffrey. *Computability and Logic*. Cambridge University Press, 2007.

Gerhard Brewka, Jürgen Dix, and Kurt Konolige. *Nonmonotonic Reasoning: An Overview*, volume 73. CSLI publications Stanford, 1997.

Wilfried Buchholz. Explaining Gentzen's consistency proof within infinitary proof theory. In *Kurt Gödel Colloquium on Computational Logic and Proof Theory*, pages 4–17. Springer, 1997.

Andrea Cantini. A theory of formal truth arithmetically equivalent to ID_1. *The Journal of Symbolic Logic*, 55(1):244–259, 1990.

Walter Carnielli, Marcelo E Coniglio, and Joao Marcos. Logics of formal inconsistency. In *Handbook of Philosophical Logic*, pages 1–93. Springer, 2007.

Walter Alexandre Carnielli and Marcelo Esteban Coniglio. *Paraconsistent Logic: Consistency, Contradiction and Negation*. Springer, 2016.

Cezary Cieslinski. *The Epistemic Lightness of Truth: Deflationism and its Logic*. Cambridge University Press, 2017.

Pablo Cobreros, Paul Egré, David Ripley, and Robert van Rooij. Tolerant, classical, strict. *Journal of Philosophical Logic*, 41(2):347–385, 2012.

Roy T Cook. There is no paradox of logical validity. *Logica Universalis*, 8(3-4): 447–467, 2014.

Marcel Crabbé. *Non-normalisation de ZF/Counterexample to normalisation for ZF*. Sketch of the result presented at the Proof Theory Symposium held in Kiel in 1974, n.d. URL `http://www.logic-center.be/Publications/Bibliotheque/\contreexemple.pdf`.

Richard L. Epstein. Relatedness and implication. *Philosophical Studies*, 36(2): 137–173, 1979.

Solomon Feferman. Toward useful type-free theories. i. *Journal of Symbolic Logic*, 49(1):75–111, 1984.

Solomon Feferman. Reflecting on incompleteness. *The Journal of Symbolic Logic*, 56(1):1–49, 1991.

Hartry Field. *Saving Truth From Paradox*. Oxford University Press, 2008.

Martin Fischer, Volker Halbach, Jönne Kriener, and Johannes Stern. Axiomatizing semantic theories of truth? *The Review of Symbolic Logic*, 8(2):257–278, 2015.

Andreas Fjellstad. Naive modus ponens and failure of transitivity. *Journal of Philosophical Logic*, 45(1):65–72, 2016a.

Andreas Fjellstad. Omega-inconsistency without cuts and nonstandard models. *The Australasian Journal of Logic*, 13(5), 2016b.

Gerhard Gentzen. *The Collected Papers of Gerhard Gentzen*. Amsterdam: North-Holland Pub. Co., 1969.

Viacheslav Nikolaevich Grišin. Predicate and set-theoretic calculi based on logic without contractions. *Mathematics of the USSR-Izvestiya*, 18(1):41, 1982.

Volker Halbach. *Axiomatic Theories of Truth*. Cambridge University Press, 2011.

Volker Halbach. *Axiomatic Theories of Truth*. Cambridge University Press, 2014.

Volker Halbach and Leon Horsten. Axiomatizing Kripke's theory of truth. *Journal of Symbolic Logic*, 71(2):677–712, 2006.

Volker Halbach and Carlo Nicolai. On the costs of nonclassical logic. *Journal of Philosophical Logic*, 47(2):227–257, 2018.

Ole Hjortland. Anti-exceptionalism about logic. *Philosophical Studies*, 174(3): 631–658, 2017.

Leon Horsten and Volker Halbach. Norms for theories of reflexive truth. In Kentaro Fujimoto, José Martínez Fernández, Henri Galinon, and Theodora Achourioti, editors, *Unifying the Philosophy of Truth*, pages 263–280. Springer Verlag, 2015.

Paul Horwich. *Truth*. Clarendon Press, 1998.

Michael N. Keas. Systematizing the theoretical virtues. *Synthese*, 195(6):2761–2793, 2018.

Georg Kreisel and Azriel Lévy. Reflection principles and their use for establishing the complexity of axiomatic systems. *Mathematical Logic Quarterly*, 14(7-12): 97–142, 1968.

Saul Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72(19):690–716, 1975.

Hannes Leitgeb. What theories of truth should be like (but cannot be). *Philosophy Compass*, 2(2):276–290, 2007.

C. Lewy, J. Watling, and P. T. Geach. Symposium: Entailment. *Aristotelian Society Supplementary Volume*, 32(1):123–172, 1958.

Vann McGee. How truthlike can a predicate be? A negative result. *Journal of Philosophical Logic*, 14(4):399–410, 1985.

Vann McGee. *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth*. Hackett Publishing, 1990.

Vann McGee. Maximal consistent sets of instances of Tarski's schema (T). *Journal of Philosophical Logic*, 21(3):235–241, 1992.

Sara Negri and Jan Von Plato. *Proof Analysis: A Contribution to HIlbert's Last Problem*. Cambridge University Press, 2011.

Sara Negri, Jan Von Plato, and Aarne Ranta. *Structural Proof Theory*. Cambridge University Press, 2008.

Francesco Paoli. *Substructural Logics: A Primer*. Springer Science & Business Media, 2013.

Lavinia Picollo. Reference in arithmetic. *The Review of Symbolic Logic*, 11(3): 573–603, 2018a.

Lavinia Picollo. Truth in a logic of formal inconsistency: how classical can it get? *Logic Journal of the IGPL*, 2018b.

Lavinia Picollo and Thomas Schindler. Deflationism and the function of truth. *Philosophical Perspectives*, 2019.

Dag Prawitz. *Natural Deduction: A Proof-Theoretical Study*. Dover Publications, 1965.

Dag Prawitz. On the idea of a general proof theory. *Synthese*, 27(1-2):63–77, 1974.

W. V. O. Quine. *Philosophy of Logic*. Harvard University Press, 1970.

Stephen Read. Anti-exceptionalism about logic. *Australasian Journal of Logic*, 2019.

William N Reinhardt. Some remarks on extending and interpreting theories with a partial predicate for truth. *Journal of Philosophical Logic*, 15(2):219–251, 1986.

Greg Restall. Multiple conclusions. In Petr Hajek, Luis Vald'es-Villanueva, and Dag Westerstaahl, editors, *Logic, Methodology and Philosophy of Science*. College Publications, 2005.

Greg Restall. Truth values and proof theory. *Studia Logica*, 92(2):241–264, 2009.

Greg Restall. Assertion, denial and non-classical theories. In Francesco Berto, Edwin Mares, Koji Tanaka, and Francesco Paoli, editors, *Paraconsistency: Logic and Applications*, pages 81–99. Springer, 2013.

David Ripley. Paradoxes and failures of cut. *Australasian Journal of Philosophy*, 91(1):139–164, 2013a.

David Ripley. Revising up: Strengthening classical logic in the face of paradox. *Philosophers' Imprint*, 13, 2013b.

David Ripley. Naive set theory and nontransitive logic. *Review of Symbolic Logic*, 8(3):553–571, 2015a.

David Ripley. Comparing substructural theories of truth. *Ergo, an Open Access Journal of Philosophy*, 2, 2015b.

Ian Rumfitt. Yes and no. *Mind*, 109(436):781–823, 2000.

Richard Mark Sainsbury. *Paradoxes*. Cambridge University Press, 2009.

Kevin Scharp. *Replacing Truth*. Oxford University Press UK, 2013.

Thomas Schindler. Some notes on truths and comprehension. *Journal of Philosophical Logic*, 47(3):449–479, 2018.

Peter Schroeder-Heister. On the notion of assumption in logical systems. In *Selected papers contributed to the sections of GAP5*, pages 22–26. Mentis, 2003.

Peter Schroeder-Heister. Paradoxes and structural rules. *Insolubles and consequences: essays in honour of Stephen Read*, pages 203–211, 2012a.

Peter Schroeder-Heister. Proof-theoretic semantics, self-contradiction, and the format of deductive reasoning. *Topoi*, 31(1):77–85, 2012b.

Peter Schroeder-Heister. Restricting initial sequents: The trade-offs between identity, contraction and cut. In Reinhard Kahle, Thomas Strahm, and Thomas Studer, editors, *Advances in Proof Theory*, pages 339–351. Springer, 2016.

Peter Schroeder-Heister and Luca Tranchini. Ekman's paradox. *Notre Dame Journal of Formal Logic*, 58(4):567–581, 2017.

Peter Schroeder-Heister and Luca Tranchini. How to Ekman a Crabbé-Tennant. *Synthese*, 2018. doi: https://doi.org/10.1007/s11229-018-02018-3.

Kurt Schütte. *Proof Theory*, volume 225. Springer Science & Business Media, 2012.

Lionel Shapiro. Validity Curry strengthened. *Thought: A Journal of Philosophy*, 2 (2):100–107, 2013.

Michael Sheard. Truth, provability and naive criteria. In Volker Halbach and Leon Horsten, editors, *Principles of Truth*, pages 169–181. Ontos Verlag, 2013.

Annika Siders. Gentzen's consistency proof without heightlines. *Archive for Mathematical Logic*, 52(3-4):449–468, 2013.

Annika Siders. Normalization proof for Peano Arithmetic. *Archive for Mathematical Logic*, 54(7-8):921–940, 2015.

Timothy Smiley. Entailment and deducibility. In *Proceedings of the Aristotelian Society*, volume 59, pages 233–254, 1958.

Peter Smith. *An Introduction to Gödel's Theorems*. Cambridge University Press, 2013.

Gaisi Takeuti. *Proof Theory*. Elsevier, 1987.

Alfred Tarski. The concept of truth in formalized languages. In Alfred Tarski, editor, *Logic, Semantics, Metamathematics*, pages 152–278. Clarendon Press, 1956.

Neil Tennant. Proof and paradox. *Dialectica*, 36(23):265–296, 1982.

Neil Tennant. *Anti-Realism and Logic: Truth as Eternal.* Oxford University Press, 1987.

Neil Tennant. On paradox without self-reference. *Analysis*, 55(3):199–207, 1995.

Neil Tennant. A new unified account of truth and paradox. *Mind*, 124(494):571–605, 2015.

Neil Tennant. Normalizability, cut eliminability and paradox. *Synthese*, 2016. doi: https://doi.org/10.1007/s11229-016-1119-8.

Neil Tennant. *Core Logic.* Oxford University Press, 2017.

Luca Tranchini. Proof-theoretic semantics, paradoxes and the distinction between sense and denotation. *Journal of Logic and Computation*, 26(2):495–512, 2016.

Anne Sjerp Troelstra and Helmut Schwichtenberg. *Basic Proof Theory.* Cambridge University Press, 2000.

Achille C. Varzi. Complementary sentential logics. *Bulletin of the Section of Logic*, 19(4):112–116, 1990.

Achille C. Varzi. Complementary logics for classical propositional languages. *Kriterion - Journal of Philosophy*, 4(1):20–24, 1992.

Jan von Plato. Translations from natural deduction to sequent calculus. *Mathematical Logic Quarterly: Mathematical Logic Quarterly*, 49(5):435–443, 2003.

Alan Weir. Naive set theory, paraconsistency and indeterminacy: Part i. *Logique Et Analyse*, 41:219–66, 1998.

Alan Weir. Naive set theory, paraconsistency and indeterminacy: Part ii. *Logique et Analyse*, 42(167/168):283–340, 1999.

Alan Weir. Naive truth and sophisticated logic. In J. C. Beall and Bradley Armour-Garb, editors, *Deflationism and Paradox.* Clarendon Press, 2005.

Alan Weir. A robust non-transitive logic. *Topoi*, 34(1):1–9, 2015.

Timothy Williamson. Semantic paradoxes and abductive methodology. In Bradley Armour-Garb, editor, *Reflections on the Liar*, pages 325–346. Oxford University Press Oxford, 2017.

Timothy Williamson. Alternative logics and applied mathematics. *Philosophical Issues*, 28(1):399–424, 2018.

Stephen Yablo. Paradox without self-reference. *Analysis*, 53(4):251, 1993.

Elia Zardini. Truth without contra(di)ction. *Review of Symbolic Logic*, 4(4):498–535, 2011.

Elia Zardini. Naive modus ponens. *Journal of Philosophical Logic*, 42(4):575–593, 2013.