A Note on the Unprovability of Consistency in Formal Theories of Truth

Kevin Davey

Abstract

Why is it that even strong formal theories of truth fail to prove their own consistency? Although Field [3] has addressed this question for many theories of truth, I will argue that there is an important and attractive class of theories of truth that he has omitted in his analysis. Such theories cannot prove that all their axioms are true, though unlike many of the cases Field considers, they do not prove that any of their axioms are false or that any of their rules of inference are not truth preserving. It is the fact that such theories are not finitely axiomatizable which stops them from proving their own consistency.

1 Introduction.

It is tempting to think that a sufficiently strong theory of arithmetic supplemented with a sufficiently strong theory of truth ought to be able to prove its own consistency. The intuitive argument is simple. Let such a theory be **U**, and assume that it is consistent. The theory **U** should be able to prove that all its axioms are true. Because the standard rules of inference are truth preserving, it follows by induction (again, within **U**) that everything provable from the axioms of **U** will also be true. Assuming **U** can prove that the sentence $0 = 1$ is not true, it then follows (still within **U**) that **U** cannot prove $0 = 1$. From this it follows (still within **U**) that **U** is consistent, for if **U** were inconsistent, **U** would prove $0 = 1$ (and would in fact prove anything). All this reasoning is conducted in **U**, and in this way, **U** should be able to prove its own consistency. Henceforth, we call this the *intuitive consistency argument.*

The result of the intuitive consistency argument, however, contradicts Gödel's Second Incompleteness Theorem. Where then does the intuitive consistency argument break down? This depends very much on the details of **U**, and Field [3] has given a truly impressive catalog of the ways in which the argument breaks down for a great number of different candidate theories **U**. However, I will argue that Field has overlooked a particularly natural (albeit underexplored) class of candidates for **U**, leading him to dismiss an otherwise important diagnosis of what goes wrong with the intuitive consistency argument. It will be part of the job of this paper to remedy this omission in Field's otherwise rich and stimulating paper.

In §2 of this paper, I will discuss what I take to be the correct diagnosis of where the intuitive consistency argument breaks down for the special class of theories of truth that I will introduce, and will also discuss Field's dismissal of this diagnosis. In §3, I will briefly discuss what this teaches us about formal theories of truth and the expressive power of first order logic, and will remark on the consequences of this for deflationist theories of truth. Technical matters will then be dealt with and relevant proofs given in the appendices.

2 Why is Consistency Unprovable?

2.1 Field's Taxonomy.

In [3], Field also presents his own intuitive consistency argument for a sufficiently strong theory **U**, which for our purposes we regard as the same as the one presented in the previous section. He divides it into 3 steps, outlined on p. 568 of [3]. In the first step, one somehow argues:

> (1) Each axiom of **U** is true.

In the second step, one argues:

> (2) Each rule of inference of **U** preserves truth (i.e., whenever the premisses of the rule are true, so is the conclusion.)

Thus, by induction, one concludes in the the third step:

> (3) All theorems of **U** are true.

The rest of the argument proceeds as outlined in the previous section. Restricting his attention at first to theories in which the underlying logic is classical (a restriction which we shall also accept for the duration of this paper), Field concludes on pp. 571-2 of [3]:

> Then aside from one totally unattractive classical theory ... the breakdown of the Consistency Argument in classical theories always occurs either because of a failure of an individual instance of (1), or because of a failure of an individual instance of (2). And when I say here that there is a failure, I mean not just that the theory does not contain the claim, I mean that it contains its negation. That is, it is always the case either that:
>
> (A) The theory employs an axiom that the theory implies is not true, or that
>
> (B) The theory employs a rule of inference that it implies is not truth-preserving.

Field dismisses outright the idea that we ought to restrict the principle of mathematical induction, and for the duration of this paper, I shall not disagree with him on this. Even given this, however, I shall argue that the diagnosis Field gives in the above paragraph is not complete, and that there is a very natural class of theories of truth whose underlying logic is fully classical, and for which neither (A) nor (B) hold. In fact, it is party because of this that I think such theories of truth are especially attractive.

2.2 Field on Infinite Axiomatizability.

Let us consider step (1) of the intutive consistency argument, according to which we prove within **U** the sentence $\mathbf{\Omega_U}$ that states that all the axioms of **U** are true:

$$(\forall x)(\mathbf{Ax_U}(x) \supset \mathbf{T}(x)) \quad (\mathbf{\Omega_U}).$$

Here $\mathbf{Ax_U}(x)$ is an appropriately chosen predicate that states that x is a Gödel number of one of the axioms of **U**. How might we prove $\mathbf{\Omega_U}$ within **U**?

In this regard, consider the rule sometimes called **NEC** (the rule of necessitation), that allows us to infer from a proof of a closed sentence X (with no undischarged assumptions) the sentence $\mathbf{T}(\ulcorner X \urcorner)$, i.e.,

NEC: if $\vdash X$, then $\vdash \mathbf{T}(\ulcorner X \urcorner)$.[1]

Suppose that our theory **U** is finitely axiomatizable, and that it consists of the axioms $\tau_1, \tau_2, ..., \tau_n$. Under such assumptions, with the use

[1] We will later generalize this to open sentences, but for now it will suffice to formulate this rule for closed sentences X only.

of **NEC** we can prove each $\mathbf{T}(\tau_i)$, and thus can prove:

$$\mathbf{T}(\tau_1)\ \&\ \mathbf{T}(\tau_2)\ \&\ ...\ \&\ \mathbf{T}(\tau_n).$$

From this, $\mathbf{\Omega_U}$ can then be fairly straightforwardly derived.[2] So step (1) of the intuitive consistency argument is easy to carry out if **U** is finitely axiomatizable and we have access to the rule **NEC**. But if **U** is not finitely axiomatizable, then even armed with **NEC**, how are we to derive $\mathbf{\Omega_U}$? There is no obvious way to do so. And thus one might find oneself tempted by the idea that the real problem with the intuitive consistency argument is that it fails in step (1), insofar as our theory **U** will typically be infinitely axiomatized, leaving it completely unclear how $\mathbf{\Omega_U}$ is to be derived.

Field considers this diagnosis of the problem with the intuitive consistency argument in section 3 of [3], but concludes there that it is mistaken, arguing that '*this ... diagnosis cannot be correct in general (or even, for all theories of sufficient strength)*' (p. 571). His argument is schematic, but may be spelled out easily as follows.

In addition to **NEC**, let us arm ourselves with the rule **CONEC** that, given any closed sentence X, allows to infer from a proof of $\mathbf{T}(\ulcorner X \urcorner)$ (with no undischarged assumptions) the sentence X, i.e.,

CONEC: if $\vdash \mathbf{T}(\ulcorner X \urcorner)$, then $\vdash X$.

Now, suppose we have some recursively axiomatized theory **U** extending a sufficiently strong fragment of arithmetic. Let the axioms of **U** be given by some (recursive) sequence $\tau_1, \tau_2, ...$ Then consider the theory $\mathbf{U}^*$ consisting of the single axiom:

$$\mathbf{F}\ \&\ (\forall x)(\mathbf{Ax_U}(x) \supset \mathbf{T}(x)) \qquad (\mathbf{U}^*)$$

where (i) $\mathbf{Ax_U}(x)$ is an appropriately chosen predicate that holds of x iff x is the Gödel number of one of the τ_i, (ii) **F** is the conjunction of a finite number of the axioms of **U**, sufficiently strong that **F** proves $\mathbf{Ax_U}(n)$ whenever n is actually the Gödel number of an axiom of **U**.[3]

[2]It suffices to assume that $\mathbf{Ax_U}$ has been chosen so that

$$(\forall x)[\mathbf{Ax_U}(x) \leftrightarrow (x = \ulcorner \tau_1 \urcorner \vee ... \vee x = \ulcorner \tau_n \urcorner)]$$

can be proven in **U**.

[3]For example, if we suppose that the predicate $\mathbf{Ax_U}(x)$ is a straightforward formulation of the claim that a certain Turing machine halts on input x, then for any n for which this Turing machine does indeed halt on n, $\mathbf{Ax_U}(n)$ will be provable from the axioms of Robinson Arithmetic. Thus, we can let **F** be the conjuction of the axioms of Robinson Arithmetic.

The theory $\mathbf{U}^*$ thus proves each $\mathbf{Ax_U}(\ulcorner \tau_i \urcorner)$ (because each such sentence follows from $\mathbf{F}$), and thus $\mathbf{U}^*$ also proves each $\mathbf{T}(\ulcorner \tau_i \urcorner)$. Armed with the rule **CONEC**, $\mathbf{U}^*$ then proves each τ_i. Consequently, $\mathbf{U}^*$ proves all the theorems that **U** proves, as it proves all of the axioms of **U**. So if **U** is a sufficiently strong theory of truth, then so will be $\mathbf{U}^*$ (and indeed, it may be even stronger than **U**.) However, $\mathbf{U}^*$ is finitely axiomatized. Thus, Field concludes that it cannot be the case that the *general* problem with the intuitive consistency argument is that theories of truth are infinitely axiomatizable, and thus that step (1) of the intuitive consistency argument cannot be completed. Rather, the general problem with the intuitive consistency argument must lie elsewhere.

Field goes on to claim, as we have seen, that for all but a few 'totally unattractive' candidates for **U** the problem lies in the fact that **U** either employs an axiom that the theory itself implies is not true (case (A) in the quote given in the previous section), or that the theory employs a rule of inference that the theory itself implies is not truth preserving (case (B)). Thus, 'totally unattractive' theories aside, Field's claim is that even when step (1) of the intuitive consistency argument fails, it fails not because of the infinite axiomatizability of the theory in question, but rather because the theory proves that one of its axioms is not true.

However, the import of Field's argument here is less than completely clear. Of course, in *some* sense part of what Field says here is right - infinite axiomatizability cannot be the *fully general* reason for the breakdown of the intuitive consistency argument, because some theories of truth are finitely axiomatized. But this is an exceedingly modest claim. We are, after all, not interested in just any old theories of truth, but rather in *attractive* theories of truth (whatever those may be.) If we do not restrict ourselves to attractive theories of truth, then there are of course all sorts of odd ways in which the intuitive consistency argument can break down, and indeed, by such standards it is not really clear that there *is* any such thing as 'the fully general reason' for the breakdown of the intuitive consistency argument. (Even Field, after all, must exclude what he deems to be 'totally unattractive' theories from his final analysis and taxonomy.) However, if we restrict ourselves to more attractive theories of truth, then perhaps we *can* give a general reason for the breakdown of the intuitive consistency argument, and presumably this is the spirit in which we are to understand Field as trying to proceed. The outcome of such a project will, however, obviously depend on what one's standards of 'attractiveness' in a theory of truth are.[4]

Field, however, arrives at his conclusion not by giving some sort of proof that certain criteria for an attractive theory of truth entail either (A) or (B). Rather, he arrives at his result by surveying what he views as reasonable contenders for our most attractive theories of truth, and noting that (A) or (B) hold of all of them. However, I shall argue that for a particularly attractive and natural class of theories of truth that Field has overlooked, it is precisely their infinite axiomatizability that blocks step (1) of the intuitive consistency argument, and that it is there and only there that the intuitive consistency argument fails for such theories. In particular, Field's diagnoses (A) and (B) do not hold of such theories - and indeed, this fact only makes such theories of truth even more attractive. Field's conclusion that attractive theories of truth fail to prove their own consistency due to reasons (A) or (B) is therefore not correct.

2.3 Minimal theories of truth.

2.3.1 Defining MT.

I maintain that Field's diagnosis overlooks a very natural and compelling class of theories of truth. These theories all extend a specific minimal theory of truth which I define in this section.

This minimal theory of truth is an extension of **PA** whose signature contains $0, ', +, \cdot$, as well as a unary predicate **T** (a truth predicate).[5] This minimal theory also contains additional axioms stating that classical, first-order logical inference preserves truth. For example, we can state that the rule of inference modus ponens is truth preserving as follows:

$$(\forall x, y)[[\text{Form}(x) \,\&\, \text{Form}(y)] \supset [(\mathbf{T}(x \supset y) \;\&\; \mathbf{T}(x)) \supset \mathbf{T}(y)]]$$

where here the variables x and y range over the natural numbers, $\text{Form}(z)$ is a primitive recursive predicate stating that z is the Gödel

[4]It is worth noting that Field offers no argument that the theory $\mathbf{U}^*$ earlier defined is 'attractive'. Surely not every theory $\mathbf{V}'$ extending an attractive theory $\mathbf{V}$ is automatically attractive - in particular, whether $\mathbf{V}'$ remains attractive will depend on our view of the principles it contains that are not in $\mathbf{V}$.

[5]A weaker fragment of **PA** would suffice, but for simplicity we work with **PA**.

number of a (possibly open) sentence - i.e., a formula - and the expression $x \supset y$ that occurs inside $\mathbf{T}(\cdot)$ is to be thought of roughly as a primitive recursive function taking the Gödel numbers x and y of two expressions into the Gödel number of the corresponding conditional (though see section §2.3.2 for a more precise discussion.)

One can formalize the claim that other rules of inference preserve truth in a similar way. For simplicity, we focus here on Hilbert style formalizations of first-order logic. (For an introduction to such formalizations of first-order logic, see §2.4 of [12], or §1 of [1].) Such systems require large sets of logical axioms to compensate for their small set of rules of inference, and so we will need axioms that state the truth of these logical axioms. For example, consider the logical axiom scheme:

$$A \supset (B \supset A)$$

using the conventions above, we state the general truth of this axiom schema as follows:

$$(\forall x, y)[[\mathrm{Form}(x) \,\&\, \mathrm{Form}(y)] \supset \mathbf{T}(x \supset (x \supset y))].$$

(Again, see §2.3.2 for a discussion of how to understand expressions such as $x \supset (y \supset x)$, where x and y are Gödel numbers of formulae.)

The propositional fragment of (one version of) the Hilbert formulation of first-order logic has the axioms:

$$\begin{array}{cc}
A \supset (B \supset A), & (A \supset B) \supset ((A \supset \neg B) \supset \neg A) \\
(A \supset B) \supset ((A \supset (B \supset C)) \supset (A \supset C)), & (\neg\neg A) \supset A \\
A \supset (A \vee B), & (A \& B) \supset A \\
B \supset (A \vee B), & (A \& B) \supset B \\
((A \supset C) \& (B \supset C)) \supset ((A \vee B) \supset C), & A \supset (B \supset (A \& B))
\end{array}$$

along with the rule of modus ponens. The truth of each of these logical axiom schemes along with the claim that modus ponens preserves truth may be formalized in the way just described.

The predicate fragment of the Hibert formalization of first-order logic has the additional logical axiom schema:

$$A[x/t] \supset (\exists x)A(x), \quad (\forall x)A(x) \supset A[x/t]$$

as well as the rules of inference:

$$\frac{C \supset A}{C \supset (\forall z)A[x/z]} \quad \frac{A \supset C}{(\exists z)A[x/z] \supset C}$$

where $A(x)$ is any formula, x and z are any variables, t any term, $A[x/u]$ is the result of substituting u for every free ocurrence of x in A, and in the two rules of inference, x is a variable which does not appear free in C.[6] [7]

The truth of these new logical axiom schemes may also be formulated straightforwardly:

$$(\forall a, x, t)[[\mathrm{Form}(a) \,\&\, \mathrm{Var}(x) \,\&\, \mathrm{Term}(t)] \supset \mathbf{T}(a[x/t] \supset (\exists x)a(x))]$$

$$(\forall a, x, t)[[\mathrm{Form}(a) \,\&\, \mathrm{Var}(x) \,\&\, \mathrm{Term}(t)] \supset \mathbf{T}((\forall x)a(x) \supset a[x/t])]$$

where $\mathrm{Var}(z)$ and $\mathrm{Term}(z)$ are unary primitive recursive relations expressing the fact that z is the Gödel number of a variable or term. The claim that these new rules of inference preserve truth may also be spelt out as follows:

$$(\forall c, a, x, z)[[\mathrm{Form}(c) \& \mathrm{Form}(a) \& \ \mathrm{Var}(x) \& \ \mathrm{Var}(z) \& \ \mathrm{NotFree}(x, c)] \supset \\ [\mathbf{T}(c \supset a) \supset \mathbf{T}(c \supset (\forall z)a[x/z])]]$$

$$(\forall c, a, x, z)[[\mathrm{Form}(c) \& \mathrm{Form}(a) \& \ \mathrm{Var}(x) \& \ \mathrm{Var}(z) \& \ \mathrm{NotFree}(x, c)] \supset \\ [\mathbf{T}(a \supset c) \supset \mathbf{T}((\exists z)a[x/z] \supset c)]]$$

where $\mathrm{NotFree}(x, c)$ is a (primitive recursive) predicate which holds of the Gödel number x of a variable and the Gödel number c of a formula iff the variable in question does not appear free in the formula in question.

In this way, we may adjoin a finite set of axioms to a base theory of arithmetic and capture the idea that logical inference preserves truth.

There are of course many different complete formulations of first-order logic within a Hilbert style framework, with possibly larger sets of rules of inference or differently formulated sets of logical axioms. Which we pick is irrelevant, as the axioms stating that the logical axiom schema are true and that the rules of inference preserve truth will all be mutually derivable. For specificity, however, we work with the system just outlined for the remainder of the paper.[8]

[6] Note that terms include expressions appropriately built not just from constants but also from variables.

[7] These are the axioms presented in sections 1.1 and 2.2 of [1].

[8] There are of course conceptually more complex, non-Hilbert style formalizations of first-order logic, such as the Fitch style formulation in which there is the machinery of

Let **MT** ('minimal truth') be the theory of truth that consists of the axioms of **PA**, with the schema of induction ranging over formulae in the extended language including the truth predicate, plus the set of axioms just formulated stating that classical first-order inference preserves truth. We also include the axiom:

$$\neg\mathbf{T}(\ulcorner 0 = 1 \urcorner).$$

The axioms of **MT** are listed explicitly again in Appendix 1. It may be shown that **MT** is a subsystem of such well-studied theories of truth as **FS** and **KF** (see [8] for definitions of these systems.) As such, it is consistent.[9]

2.3.2 Some technicalities.

There are two technical points related to the proper formulation of **MT** that we must now consider.

(i) Given a formula $\theta(x_1, ..., x_i)$ with free variables $x_1, ..., x_i$, we will *not* use the notation $\mathbf{T}(\theta(x_1, ..., x_i))$ to refer to an expression of the form $\mathbf{T}(c)$, where c is the Gödel number of the formula $\theta(x_1, ..., x_i)$. (Note that such an expression $\mathbf{T}(c)$ has no free variables.) Rather, $\mathbf{T}(\theta(x_1, ..., x_i))$ will be an expression with the same free variables $x_1, ..., x_i$ denoting the relation that holds when $\mathbf{T}(\ulcorner\theta(\overline{x_1}, ..., \overline{x_i})\urcorner)$, where here $\overline{n}$ is a canonical numerical term for the number n (e.g., $0^{\prime\prime\cdots\prime}$, with n $'$s.) Put simply, in the expression $\mathbf{T}(\theta(x_1, ..., x_i))$, the argument of $\mathbf{T}(\cdot)$ is a term denoting the primitive recursive function taking $x_1, ..., x_i$ to $\ulcorner\theta(\overline{x_1}, ..., \overline{x_i})\urcorner$. We could indicate this by writing something like $\mathbf{T}(\ulcorner\theta(\overline{x_1}, ..., \overline{x_i})\urcorner)[x_1, ..., x_i]$, but this is cumbersome and so we opt for $\mathbf{T}(\theta(x_1, ..., x_i))$ instead.

By contrast, the notation $\mathbf{T}(\ulcorner\theta\urcorner)$ will be used to denote an expression of the form $\mathbf{T}(c)$ with no free variables, where c is the Gödel

assumptions which can be discharged, or the Gentzen formulation in which a set of assumptions can have multiple conclusions. Capturing the idea that these formalizations of first-order logical inference preserve truth involves complications that we will not consider here. In fact, depending on various decisions that must be made in formulating the idea that logical inference preserves truth within such frameworks, one can end up with a set of axioms of truth *stronger* than those just developed for Hilbert style systems. This shows interesting conceptual differences between these various frameworks well worthy of further attention. Nevertheless, we will sidestep these issues in the present paper by sticking exclusively to Hilbert style formalizations of classical, first-order logic when formalizing the idea that logical inference preserves truth.

[9]In fact, it may furthermore be shown that **MT** is ω-consistent.

number of the expression θ. We will actually have little occasion to use this notation. In the case of a closed formula θ we have $\mathbf{T}(\ulcorner\theta\urcorner) = \mathbf{T}(\theta)$, but otherwise $\mathbf{T}(\ulcorner\theta\urcorner)$ and $\mathbf{T}(\theta)$ are objects of different types.

As an example, suppose x and y are Gödel numbers of formulae X and Y. When we write $\mathbf{T}(x \supset (y \supset x))]$, by $x \supset (y \supset x))$ we mean the (primitive recursive) function that takes the free variables $x_1, ..., x_i$ of X and Y to $\ulcorner(X \supset (Y \supset X))(\overline{x_1}, ..., \overline{x_i})\urcorner$. This may in turn be viewed as a primitive recursive function of x, y and $x_1, ..., x_i$. The same remarks obviously apply to any way of forming compound expressions from simpler expressions. So for example, in spelling out the axioms for **MT**, when we write $\mathbf{T}(a[x/t] \supset (\exists x)a(x))$, by $a[x/t] \supset (\exists x)a(x)$ we mean the primitive recursive function of a, t, x and any free variables $x_1, ..., x_i$ of $A[X/\tau] \supset (\exists X)A(X)$ with output $\ulcorner(A[X/\tau] \supset (\exists X)A(X))(\overline{x_1}, ..., \overline{x_i})\urcorner$, where a is the Gödel number of the expression A, t is the Gödel number of the term τ, and x the Gödel number of the variable X.

(ii) There is one slight modification we make to all this which will make things a bit easier for us in the proofs that appear in Appendix 2. When we say that something like Form(z) is a primitive recursive predicate, we mean of course that Form(z) may be defined in terms of some primitive recursive function form(z) via the equation Form(z) $\leftrightarrow$ form(z) $= 1$. Thus, wherever we have written Form(z), we will understand this as an abbreviation for the expression form(z) $= 1$, where form(z) is some appropriately chosen primitive recursive function. To make things easier, we shall actually include function symbols such as form(z) in the signature of the language, adding axioms that define form(z) in terms of simpler primitive recursive functions which we also take to be in the signature of the language, and so on. This of course is precisely the same as the way in which multiplication is a primitive recursive function in the signature of the language of arithmetic defined in terms of simpler primitive recursive functions also in the signature of the language of arithmetic (namely, addition and successor.) Just like the equations that recursively define multiplication in terms of addition and successor, the general equations that define a primitive recursive function in terms of simpler primitive recursive functions may be written as universally quantified equations (or simply as equations with free variables, if we are willing to have open sentences amongst our axioms.[10]) In this way, it is actually easier for us to work with the extension of **MT** obtained by first adding certain

[10]For example, if a primitive recursive function f is defined by primitive recursion over

primitive recursive functions to the signature of the language, and then adding the defining axioms of such primitive recursive functions to the axioms of **MT**.

We will take this approach with all the 'syntactic' predicates and functions we introduce (all of which are primitive recursive), such as Var(x), NotFree(x, c), $a[x/t] \supset (\exists x)a(x)$ and so on (all understood as primitive recursive functions, as explained in (i)), along with all the primitive recursive functions on which the definitions of such functions depend. The set of such primitive recursive functions we shall need to use in this way is large, but finite. We extend the signature of the language of arithmetic to include all such function symbols, and add the defining equations for such primitive recursive functions to our axioms, also allowing induction over predicates involving such new function symbols (and the truth predicate). We finally add the set of axioms stating that classical first-order inference (in this extended language) preserves truth, as well as the sentence $\neg\mathbf{T}(\ulcorner 0 = 1 \urcorner)$. Call the resulting system $\mathbf{MT}^+$. The system $\mathbf{MT}^+$ can be shown to be a conservative extension of **MT**.

For simplicity, we will state the main results in what follows in terms of **MT**, though in the appendix it will be easier to work in terms of $\mathbf{MT}^+$. Explicit lists of the axioms of all relevant axiom systems are included in Appendix 1, and in Appendix 2 we discuss how the relevant theorems about $\mathbf{MT}^+$ entail theorems about **MT**.[11]

2.3.3 MT and the Intuitive Consistency Argument.

These two technical remarks having been made, we now return to the main argument. We will be interested in theories that extend **MT**, and which employ only some complete and sound set of rules for first order logic, with no other rules of inference. Call a theory which employs only some set of complete and sound rules for first order logic *purely logical.* So for example, any theory of truth which employs

primitive recursive functions g and h, its defining equations will have the form:

$(\forall \vec{y})(f(0, \vec{y}) = g(\vec{y}))$

$(\forall \vec{y})(f(x', \vec{y}) = h(x, f(x, \vec{y}), \vec{y}))$.

Primitive recursive functions defined in terms of simpler primitive recursive functions by composition may also be defined by equations of the same form.

[11]The main advantage of working with $\mathbf{MT}^+$ is that it allows us to think of predicates like Form(z) as atomic, which will make proof theoretic analysis easier.

the rules **NEC** or **CONEC** (not just as derived rules, but as basic rules) is *not* purely logical. Our main interest will be in purely logical theories of a certain sort extending **MT**.

The important thing to note about purely logical theories of truth extending **MT** is that insofar as **MT** explicitly states that logical inference preserves truth, step (2) of the intuitive consistency is guaranteed to work for them. If step (1) of the intuitive consistency argument can be made to work for some such theory **U**, this would mean that **U** proves that all its theorems are true:

$$\mathbf{U} \vdash (\forall x)(\mathbf{Pr_U}(x) \supset \mathbf{T}(x)).$$

Thus we would have:

$$\mathbf{U} \vdash \mathbf{Pr_U}(\ulcorner 0 = 1 \urcorner) \supset \mathbf{T}(\ulcorner 0 = 1 \urcorner),$$

and because **U** contains the axiom $\neg\mathbf{T}(\ulcorner 0 = 1 \urcorner)$, we would then have:

$$\mathbf{U} \vdash \neg\mathbf{Pr_U}(\ulcorner 0 = 1 \urcorner).$$

Under the assumption that step (1) works for some such **U**, we therefore have that **U** proves its own consistency. Thus, for a purely logical theory of truth extending **MT**, the only place the intuitive consistency argument can go wrong is in step (1). We will consider exactly how for an important class of such theories in what follows.

2.4 On NEC and CONEC.

The rules of inference **NEC** and **CONEC** capture the idea that for any sentence X, we can infer $\mathbf{T}(\ulcorner X \urcorner)$ from X, and vice-versa. As such, they are indispensable in any serious theory of truth. Of course, such inferences could be made without introducing any new rules of inference if we had the Tarski biconditionals $X \leftrightarrow \mathbf{T}(\ulcorner X \urcorner)$ in our theory (either as axioms or as theorems). But under minimal assumptions Tarski biconditionals quickly lead to contradictions, whereas rules like **NEC** and **CONEC** are a bit better behaved. This has made rules of inference like **NEC** and **CONEC** popular in formal theories of truth[12], especially since the pioneering paper of Friedman and Sheard

[12]The question of whether **NEC** and **CONEC** are *sufficient* to generate all facts about truth in which we might be interested is a question we do not go into here. One worry is that with such rules alone, we cannot prove conditional claims such as 'if everything John says is true, then if John said it will rain tomorrow, then it will rain tomorrow.'

[5] in which such theories were first systematically studied in detail.[13] In light of all this, one might think that purely logical theories of truth, insofar as they do *not* have **NEC** and **CONEC** among their basic rules, cannot be of much interest.

But such a conclusion would be premature. Even in a purely logical theory, **NEC** and **CONEC** could well turn out to be admissible as *derived rules.* Let us begin by considering the rule **NEC**. The idea that having proved X we have also proved $\mathbf{T}(\ulcorner X \urcorner)$ is very compelling. It seems to me, however, that it is not 'axiomatically' compelling, but rather compelling in virtue of a piece of reasoning we feel confident in carrying out. In particular, suppose we have proved X from the axioms $\tau_1, ...\tau_i$ of some purely logical theory **U** extending **MT**. In virtue of being committed to the axioms $\tau_1, ...\tau_i$, we are surely committed to the truth of these axioms - i.e., to $\mathbf{T}(\tau_1), ...\mathbf{T}(\tau_i)$. But if from $\tau_1, ...\tau_i$ we can prove X, then from $\mathbf{T}(\tau_1), ...\mathbf{T}(\tau_i)$ and the idea that logical inference preserves truth, we should be able to prove $\mathbf{T}(\ulcorner X \urcorner)$.

What this means is that so long our purely logical theory **U** extending **MT** is such that whenever τ is an axiom of **U**, $\mathbf{T}(\tau)$ is also an axiom of **U**, then **NEC** will be a *derived rule* of **U** - that is, while not one of the basic rules of inference permissible in **U**, it will nevertheless take provable claims to provable claims.

To state this fact clearly, we introduce some notation. For any axiomatizable theory **U**, let $\mathbf{U}^\circ$ be the theory whose axioms consist of every sentence of the form $\mathbf{T}^{(n)}(\tau)$ (with $n \geq 0$), where τ is one of the axioms of **U** (logical or otherwise.) Here $\mathbf{T}^{(n)}(\tau)$ is $\mathbf{TT}...\mathbf{T}(\tau)$ with n **T**s, and $\mathbf{T}^{(0)}(\tau)$ is τ.) We then have the following easy theorem:

Theorem 1. *Let* **U** *be a purely logical theory extending* **MT**. *Then if* $\mathbf{U}^\circ \vdash X$, *then* $\mathbf{U}^\circ \vdash \mathbf{T}(\ulcorner X \urcorner)$.

Proof. Under the stated assumptions about **U**, let us suppose that $\mathbf{U}^\circ \vdash X$. Suppose the proof of X in $\mathbf{U}^\circ$ is written out in the Hilbert style formalism previously discussed. The proof that $\mathbf{U}^\circ \vdash \mathbf{T}(\ulcorner X \urcorner)$ is by induction on the length of such a proof.

For the base case, note that if X is an axiom (logical or otherwise) of $\mathbf{U}^\circ$, then we immediately have that $\mathbf{U}^\circ \vdash \mathbf{T}(\ulcorner X \urcorner)$ by definition of $\mathbf{U}^\circ$. For the inductive step, suppose that the last step of the proof of X

[13] Inspired by the Fitch formalism for first-order logic, there is perhaps also the idea that predicates and operators in our language must somehow be associated with some sort of introduction and elimination rules that give them content. The rules **NEC** and **CONEC** look like they fulfill this role nicely.

is an application of modus ponens from sentences Y and $Y \supset X$ that have been proven in $\mathbf{U}^\circ$. By inductive hypothesis, we may suppose that $\mathbf{U}^\circ \vdash \mathbf{T}(\ulcorner Y \urcorner)$ and $\mathbf{U}^\circ \vdash \mathbf{T}(\ulcorner Y \supset X \urcorner)$. Because **U** extends **PA**, we may suppose that $\text{Form}(\ulcorner Y \urcorner)$ and $\text{Form}(\ulcorner Y \supset X \urcorner)$ are also provable in $\mathbf{U}^\circ$. From all this and the following axiom of **MT**:

$$(\forall x, y)[[\text{Form}(x) \,\&\, \text{Form}(y)] \supset [(\mathbf{T}(x \supset y) \;\&\; \mathbf{T}(x)) \supset \mathbf{T}(y)]]$$

it then follows that $\mathbf{U}^\circ \vdash \mathbf{T}(\ulcorner X \urcorner)$. The case in which the last step of the proof of X is one of the other basic rules of inference may be dealt with similarly. □

What this theorem shows is that even a purely classical theory of the right form (i.e., of the form $\mathbf{U}^\circ$) can have **NEC** as a derived rule, and thus perhaps be attractive as a theory of truth.

What about **CONEC**? Articulating conditions under which it can be proven that a theory of the form $\mathbf{U}^\circ$ admits **CONEC** as a derived rule is non-trivial. For certain specific theories **U**, however, such a result does in fact hold. For example:

Theorem 2. *Let* **U** *be either the theory* **MT** *or* **FSN**. *Then for any closed sentence X, if* $\mathbf{U}^\circ \vdash \mathbf{T}(\ulcorner X \urcorner)$*, then* $\mathbf{U}^\circ \vdash X$.

(The theory **FSN** is defined in Chapter 14 of Halbach [8], and we also list its axioms in Appendix 1.) The proof of this result is much more complicated and so occupies Appendix 2. That the result holds for **FSN** is a consequence of the argument Halbach gives in §14.1 of [8] for the redundancy of **CONEC** in **FS** - see in particular the discussion on pp. 174-5. In Appendix 2, we therefore focus on the argument for **MT**, giving a proof theoretic argument for this case.

What these theorems show is that even certain purely classical theories can allow **NEC** and **CONEC** as derived rules, and thus, I think, be highly attractive as theories of truth. Indeed, I think that purely classical theories of truth that allow **NEC** and **CONEC** as derived rules have virtues that other theories of truth lack. In particular, they show that **NEC** and **CONEC** are not valid rules of inference as a matter of brute fact, but rather in virtue of something about the structure of our axiomatic commitments (e.g., the closure of our axiomatic commitments under the operation that takes X to $\mathbf{T}(\ulcorner X \urcorner)$), and the character of the laws of logic.

Note that if **U** is purely logical theory extending **MT**, then so is $\mathbf{U}^\circ$. As already mentioned, this means that it must be in step (1) that the intuitive consistency argument fails for such a $\mathbf{U}^{\circ}$. Note that regardless of whether or not $\mathbf{U}$ is finitely axiomatizable, $\mathbf{U}^{\circ}$ will have infinitely many axioms. There is thus no clear strategy for proving the sentence $\mathbf{\Omega}_{\mathbf{U}^{\circ}}$ stating that all the axioms of $\mathbf{U}$ are true:

$$(\forall x)(\mathbf{Ax}_{\mathbf{U}^{\circ}}(x) \supset \mathbf{T}(x)) \qquad (\mathbf{\Omega}_{\mathbf{U}^{\circ}}).$$

Given that this is the only place the intuitive consistency argument can go wrong for such theories, Gödel's Second Incompleteness Theorem then entails that $\mathbf{U}^{\circ}$ in fact does not prove $\mathbf{\Omega}_{\mathbf{U}^{\circ}}$. Thus, contrary to what Field argues, it is indeed the non-finite axiomatizability of $\mathbf{U}^{\circ}$ that causes a breakdown in the intuitive consistency argument.

More importantly, note that neither of Field's diagnoses (A) or (B) accounts for the failure of the intuitive consistency argument in these cases. There is no axiom of $\mathbf{U}^{\circ}$ which our theory proves to be not true (assuming $\mathbf{U}^{\circ}$ is consistent), as for each axiom τ of $\mathbf{U}^{\circ}$, $\mathbf{T}(\tau)$ is also an axiom of $\mathbf{U}^{\circ}$. And by the assumption that $\mathbf{U}$ is purely classical, we know that the only basic rules of inference of $\mathbf{U}^{\circ}$ are the standard rules of first order logic. From the fact that $\mathbf{U}^{\circ}$ extends **MT** we know that $\mathbf{U}^{\circ}$ proves that each of these rules is truth preserving. So neither (A) nor (B) obtain.

In sum, purely classical theories of the form $\mathbf{U}^{\circ}$ extending **MT** are natural theories of truth for which the intuitive consistency argument breaks down merely in virtue of their non-finite axiomatizability, and not in virtue of the reasons (A) or (B) offered in Field's analysis.

2.5 Derived and Basic Rules

I have argued that for certain natural theories $\mathbf{U}$ extending **MT**, both **NEC** and **CONEC** can be treated as derived rules of $\mathbf{U}^{\circ}$. This means that within the context of such a theory $\mathbf{U}^{\circ}$, $\mathbf{T}(\ulcorner X \urcorner)$ may be inferred from X, and vice versa. Consider then the sentence stating that these derived rules of inference are truth preserving; i.e., that if the premise of such a rule is true, then so is its conclusion:

$$(\forall x)[\text{Form}(x) \supset [\mathbf{T}(x) \leftrightarrow \mathbf{TT}(x)]] \qquad (i)$$

Unfortunately, if $\mathbf{U}$ extends **MT**, $\mathbf{U}^{\circ}$ proves that (i) entails a contradiction. (For a proof, see Theorem 13.8 of [8].) Thus, $\mathbf{U}^{\circ}$ proves the negation of (i). The negation of (i) seems to say that it is *not* the case that both **NEC** and **CONEC** are truth preserving. But does this not mean that Field is right after all, and that for the sorts of theories for which Theorems 1 and 2 hold, it *is* indeed true that 'the theory employs a rule of inference that it implies is not truth-preserving'? Or using Field's terminology, is not diagnosis (B) correct after all?

Here, several points are in order. First of all, even if it is true of the theories under consideration that they admit a rule of inference that the theory implies is not truth-preserving, this does not prevent step (2) of Field's intuitive consistency argument from going through. The intuitive consistency argument only requires that the theory in question be able to prove that its *basic* rules of inference are truth preserving, and this will be true of any purely classical theory extending **MT**. The intuitive consistency argument does not require that we be able to somehow prove that every possible *derived* rule of inference is truth preserving. There are of course interesting questions about rules like **NEC** worth asking, but they are not what matter when trying to figure out where the intuitive consistency argument goes wrong in purely classical theories.

But second of all, there is a misunderstanding about the **NEC** and **CONEC** rules implicit in this worry that is actually independent of whether we think of these rules as basic or derived. Note that the rules **NEC** and **CONEC** differ from the other ordinary rules of logic in that they may not be applied arbitrarily, but only when something has been proved unconditionally. More specifically, these rules cannot be applied within subderivations. Let us focus on **NEC**, for example. In saying that this rule preserves truth, we need *not* say anything as strong as that whenever a sentence X is true, it is true that it is true. That is, we need *not* endorse:

$$(\forall x)[\mathrm{Form}(x) \supset [\mathbf{T}(x) \supset \mathbf{TT}(x)]] \qquad (ii)$$

Rather, to say that the rule **NEC** preserves truth is to make the much more modest claim that whenever a sentence X *that has been proved unconditionally* is true, then it is true that it is true, i.e.,

$$(\forall x)[\mathrm{Form}(x) \supset [\mathbf{Prov}_{\mathbf{U}^\circ}(x) \supset [\mathbf{T}(x) \supset \mathbf{TT}(x)]] \qquad (iii)$$

The corresponding claim for **CONEC** involves the converse claim, and when conjoined with (iii) gives:

$$(\forall x)[\mathrm{Form}(x) \supset [\mathbf{Prov}_{\mathbf{U}^\circ}(x) \supset [\mathbf{T}(x) \leftrightarrow \mathbf{TT}(x)]] \qquad (iv)$$

Now, while (i) is not consistent with the theories $\mathbf{U}^\circ$ in which we are interested, (iv) in general *is*. In particular, note that (iv) follows from the combination of:

$$(\forall x)[\text{Form}(x) \supset [\mathbf{Prov}_{\mathbf{U}^\circ}(x) \supset \mathbf{Prov}_{\mathbf{U}^\circ}(\mathbf{T}(x))] \qquad (v)$$

and

$$(\forall x)[\text{Form}(x) \supset [\mathbf{Prov}_{\mathbf{U}^\circ}(x) \supset \mathbf{T}(x)] \qquad (vi)$$

Claim (v) is provable in any theory in which the proof of Theorem 1 can be formalized, and thus is provable in **PA**. Claim (vi) is provable in any extension of **MT** that proves $\mathbf{\Omega}_{\mathbf{U}^\circ}$ – i.e., any extension of **MT** that proves the sentence stating that all the axioms of $\mathbf{U}^\circ$ are true. Putting these facts together means that (iv) is actually a theorem of $\mathbf{U}^\circ + \mathbf{\Omega}_{\mathbf{U}^\circ}$.

In sum, as **NEC** and **CONEC** are derived rules, questions about whether they are truth preserving have no impact on the intuitive consistency argument. Having said that, there is nothing problematic about the claim that the rules **NEC** and **CONEC** are truth preserving, so long as this claim is properly formulated.

3 Corollaries.

3.1 Commitment and Provability.

In light of the fact that purely classical theories of the form $\mathbf{U}^\circ$ extending **MT** do not prove the sentence $\mathbf{\Omega}_{\mathbf{U}^\circ}$ asserting the truth of all their axioms, one might wonder how attractive such theories of truth really are. One might think that part of the point of appending a theory of truth to **PA**, for example, is to capture the large class of sentences to which we are committed in virtue of being committed to the axioms of **PA**. The typical mathematician, for example, is not just committed to the truth of the axioms of **PA** individually, but is also committed to their collective truth, and the truth of their collective truth, and so on. One might indeed hope that the 'correct' theory of truth appended to **PA** would formalize all these commitments and make them explicit. But insofar as no purely classical theory of truth $\mathbf{U}^\circ$ can prove the sentence saying that all its axioms are true - surely something to which anyone committed to $\mathbf{U}^\circ$ should also be committed - it would seem to follow that purely classical theories of truth of the form $\mathbf{U}^\circ$ cannot be the theories of truth we were looking for.

I think, however, that this is a mistaken way to think about theories of truth. It is far from clear that there is some sort of special theory consisting of all and only those propositions to which we are committed in virtue of being committed to **PA**. Perhaps one way to make this claim precise would be to deny that the set of sentences to which we are committed in virtue of being committed to **PA** is recursively axiomatizable. Related to this, I am sometimes tempted by the idea that there is something 'open ended' about precisely what we are committed to when we are committed to a theory like **PA**, and that as a result, no theory of truth appended to **PA** can make explicit *all* the commitments that flow from our more basic commitment to **PA**. Developing such lines of argument will not, however, be the job of this paper, and so I will not press such points further here. (For an elaboration and defense of such lines of thought see Franzen [4].)

Independently of whether one is willing to buy the rhetoric of open-endedness, I think the following argument against the above conception of theories of truth can be made. Just because our commitment to some theory **X** means that we 'ought to accept' some sentence **Y**, it does not follow that **Y** must be provable from **X**. Gödel's Second Incompleteness Theorem is an obvious example of this. Suppose, for example, that **X** is some theory to which we are committed. This commitment, I think, entails a committment to **Con**(**X**). Under minimal assumptions, this will of course not be provable from **X**. And so the fact that our commitment to some theory **X** means that we are committed to some sentence **Y** does not mean that **Y** should be provable from **X** in any sense. Reflection principles are another closely related example of this phenomenon. According to Feferman [2], acceptance of a theory implicitly commits one to accept a reflection principle for that theory, according to which all the theorems of that theory are true. It does not follow, however, that such a reflection principle is provable in the original theory (as Feferman of course knew). Indeed, in general it is not.

I claim that the fact that our purely classical theory $\mathbf{U}^\circ$ does not prove the sentence saying that all its axioms are true is yet another example of this general phenomenon. Just because we are committed to the sentence $\boldsymbol{\Omega}_{\mathbf{U}^\circ}$ that says that all the axioms of $\mathbf{U}^\circ$ are true does not mean that $\boldsymbol{\Omega}_{\mathbf{U}^\circ}$ should be provable from $\mathbf{U}^\circ$, and in general it is not. I thus see no reason to think of this basic unprovability result as some sort of defect in purely classical theories of the form $\mathbf{U}^\circ$ extending **MT**.

3.2 Expressive Limitations.

What underlies much of what we have been talking about is a basic expressive limitation of first-order logic with the usual classical rules of inference. Given an infinitely axiomatized theory $\mathbf{U}^\circ$ consisting of sentences $\tau_1, \tau_2, ...$, it is tempting to think that armed with the truth predicate, we can simply assert all of these axioms at once. Consider, for example, the English sentence $(*)$:

All the τ_i are true. $(*)$

Armed with a truth predicate and a little arithmetic in the background, the formalization of this sentence poses no problem. However, within the context of purely classical theories of truth extending **MT**, what we have seen is that this sentence says *more* than just that the axioms of $\mathbf{U}^\circ$ are individually true. To say that *all* the τ_i are true is to do more than assert the τ_i individually. Indeed, $(*)$ is not even provable in $\mathbf{U}^\circ$ in general. Thus, we cannot assert all the axioms $\tau_1, \tau_2, ...$ at once without stating something that goes beyond the theory whose axioms consists of the τ_i individually. This is simply a fact about the expressive capacities of first-order logic. Moreover, it is precisely this fact that blocks the intuitive consistency argument in the case of purely classical theories of truth of the form $\mathbf{U}^\circ$ extending**MT**.

3.3 Deflationism.

It seems to me that this has consequences for disquotationalist theories of truth, at least understood a certain way. According to disquotationalist theories of truth, one of the functions of the truth predicate is to allow us to express infinitely many sentences at once. However, if the argument of the previous sections is right, in expressing infinitely many sentences at once we often go beyond the theory axiomatized by these sentences individually. More specifically, we have seen that if **U** is a purely classical theory extending **MT**, adding the sentence 'all the axioms of $\mathbf{U}^\circ$ are true' to $\mathbf{U}^\circ$ does *not* give a conservative extension of $\mathbf{U}^\circ$. In particular, it allows us to prove the sentence $\mathbf{Con}(\mathbf{U}^\circ)$, which is not itself generally provable in $\mathbf{U}^\circ$.

There is no denying that the sentence 'all the axioms of $\mathbf{U}^\circ$ are true' is indeed a sentence from which the individual axioms of $\mathbf{U}^\circ$ can be deduced, so long as one is allowed to use **CONEC**, and we have enough arithmetic in the background. If that is all it means to say that the sentence 'all the axioms of $\mathbf{U}^\circ$ are true' expresses all the axioms of $\mathbf{U}^\circ$, then I have no problem with the claim that the truth predicate allows us to express infinitely many sentences at once. But if the thought is that to express an infinite list of sentences is to make a claim which does not entail anything not already entailed by the theory consisting of the original infinite set of sentences listed individually, then I think that the truth predicate does *not* make such an act of expression possible. Whether disquotationalism demands this notion of expression is not something I will quibble over here, though I think there is a case to be made that it does. Insfoar as it does, disquotationalist claims about the function of the truth predicate cannot be correct.[14]

Appendix 1.

For reference, it will be useful to write out explicitly the axioms of $\mathbf{MT}$ and $\mathbf{MT}^+$, along with their counterparts $\mathbf{MT}^\circ$ and $(\mathbf{MT}^+)^\circ$.

Recall that the theory $\mathbf{MT}^+$ is the conservative extension of $\mathbf{MT}$ described in §2.3. The theory $\mathbf{MT}^+$ consists of the axioms:

$(\mathbf{MT1}_s)$ $s = s$

$(\mathbf{MT2}_{s_1,\dots,t_n,f})$ $(s_1 = t_1 \;\&\; \dots \;\&\; s_n = t_n) \supset f(s_1, \dots, s_n) = f(t_1, \dots, t_n)$

$(\mathbf{MT3}_{s_1,\dots,t_n,R})$ $(s_1 = t_1 \;\&\; \dots \;\&\; s_n = t_n) \supset (R(s_1, \dots, s_n) \supset R(t_1, \dots, t_n))$

$(\mathbf{MT4})$ $x' \neq 0$

$(\mathbf{MT5})$ $x' = y' \supset x = y$

$(\mathbf{MT6})$ $x + 0 = x$

$(\mathbf{MT7})$ $(x + y)' = x + y'$

$(\mathbf{MT8})$ $x.0 = 0$

$(\mathbf{MT9})$ $x.y' = x.y + x$

$(\mathbf{MT10})$ axioms giving the definitions of further primitive recursive functions

$(\mathbf{MT11}_\phi)$ $(\phi[z/0] \;\&\; (\forall x)(\phi[z/x] \supset \phi[z/x'])) \supset (\forall x)\phi[z/x]$

where in $(\mathbf{MT11}_\phi)$, we assume that x does not appear free in $\phi[z/0]$.

To make notation less cumbersome, we have presented (and will continue to present) axioms as open sentences. In the version of the

[14]This argument is reminiscent of the sort of argument found in §3 of Heck's critique of disquotationalism in [9].

Hilbert calculus we have been working with, any open sentence is interderivable with the corresponding sentence in which all free variables are universally quantified over.

In (**MT1**-**MT3**), s, the s_i, and the t_i are arbitrary terms, f any function and R any relation in the signature of the language $\mathcal{L}$. Thus, (**MT1**-**MT3**) are schemes and not single axioms. In (**MT4**-**MT11**) (and in the axioms that follow), x, y and z are arbitrary variables, which can of course be replaced with any other variable in our language. Axiom (**MT10**) includes the definitions for the large (but finite) class of primitive recursive functions needed to formalize notions such as $\text{Form}(x)$ or $\text{NotFree}(x, c)$, as discussed in §2.3. These primitive recursive definitions all have the same positive, purely equational form found in (**MT3**-**MT6**). In axiom (**MT11**), ϕ ranges over all formulae in the full language.

We furthermore have the axioms:

(MT12) $\neg\mathbf{T}(0 = 1)$

(MT13$_i$) $(\text{Form}(x_1)\ \&\ \text{Form}(x_2)\ \&\ ...) \supset \mathbf{T}(s_i(x_1, x_2, ...))$

where in **MT13**, each $s_i(x_1, x_2, ...)$ $(i = 1, ...k)$ is a primitive recursive function taking Gödel numbers of formulae to the Gödel number of a larger formulae corresponding to one of the logical axiom schemes of the Hilbert calculus as spelt out in §2.3. For example, $s_1(x_1, x_2)$ is a primitive recursive function that takes the Gödel numbers of formulae θ_1 and θ_2 to the Gödel number of the formula $\theta_1 \supset (\theta_2 \supset \theta_1)$; we write this as $s_1(x_1, x_2) = x_1 \supset (x_2 \supset x_1)$. We have one such axiom for each of the logical axiom schemas of the Hilbert calculus. (Here k is the number of such logical axiom schemes.)

In addition, we also have the axioms:

(MT14) $(\text{Form}(x_1)\ \&\ \text{Form}(x_2)) \supset ([\mathbf{T}(x_1 \supset x_2)\ \&\ \mathbf{T}(x_1)] \supset \mathbf{T}(x_2))$

(MT15) $(\text{Form}(c)\ \&\ \text{Form}(a)\ \&\ \text{Var}(x)\ \&\ \text{Var}(z)\ \&\ \text{NotFree}(x, c)) \supset$
$(\mathbf{T}(c \supset a(x)) \supset \mathbf{T}(c \supset (\forall z)a[x/z]))$

(MT16) $(\text{Form}(c)\ \&\ \text{Form}(a)\ \&\ \text{Var}(x)\ \&\ \text{Var}(z)\ \&\ \text{NotFree}(x, c)) \supset$
$(\mathbf{T}(a(x) \supset c) \vdash \mathbf{T}((\exists z)a[z/x] \supset c))$

The axioms of $\mathbf{MT}^+$ are just **MTa** for $a = 1, ...16$. The axioms of **MT** may be obtained by omitting **MT10** and placing appropriate restrictions on the language used in the terms, formulae etc. used in the remaining axioms.

For any axiom system **U**, the axioms of $\mathbf{U}^\circ$ are just the closure of the set of axioms of **U** under the operation that takes a formula ϕ

to $\mathbf{T}(\ulcorner\phi\urcorner)$. Thus, whenever ϕ is an axiom of $\mathbf{U}^\circ$, so too is $\mathbf{T}^{(n)}(\ulcorner\phi\urcorner)$ for $n > 0$, where here $\mathbf{T}^{(n)}(\ulcorner\phi\urcorner)$ is just $\mathbf{TT}...\mathbf{T}(\ulcorner\phi\urcorner)$ with n **T**s, and $\mathbf{T}^{(0)}(\ulcorner\phi\urcorner)$ is ϕ.

For each axiom (scheme) **MTa** ($a = 1, ...16$) and $n \geq 0$, we let $\mathbf{MTa}^{(n)}$ denote $\mathbf{T}^{(n)}(\ulcorner\phi\urcorner)$, where ϕ is the axiom **MTa** (or an instance thereof, if **MTa** is a schema.) With this notation in place, the axioms of $(\mathbf{MT}^+)^\circ$ are just $\mathbf{MTa}^{(n)}$ for all $n \geq 0$, and $a = 1, ...16$. The axioms of $(\mathbf{MT})^\circ$ may be defined from **MT** similarly.

Appendix 2.

*** check appendices with immodest version ***

Here we prove:

Theorem 2. *For any closed sentence X, if $\mathbf{MT}^\circ \vdash \mathbf{T}(\ulcorner X\urcorner)$, then $\mathbf{MT}^\circ \vdash X$.*

We prove Theorem 2 proof-theoretically, by repurposing in a fairly straightforward way the cut-elimination techniques found in Gentzen's argument for the consistency of arithmetic (see [6], [7].) In fact, what we will prove is:

Theorem 2′. *For any closed sentence X, if $(\mathbf{MT}^+)^\circ \vdash \mathbf{T}(\ulcorner X\urcorner)$, then $(\mathbf{MT}^+)^\circ \vdash X$.*

Here $\mathbf{MT}^+$ is the conservative extension of **MT** outlined in Appendix 1 and discussed in §2.3. Because $\mathbf{MT}^+$ is indeed a conservative extension of **MT**, Theorem 2′ is easily seen to entail Theorem 2.

The proof of Theorem 2′ will involve a detour through the sequent calculus. In the version of the sequent calculus we will be interested in, we treat the left and right side of sequents as finite *multisets* of formulae, eliminating the need for a rule of exchange. In our system, the formulae that appear in the sequents will be written in a language that extends the usual language of arithmetic insofar as it includes (i) a truth predicate **T**, and (ii) new function symbols for a large class of primitive recursive functions, as described in §2.3. We will also use the symbol $\perp$, which can be replaced by $0 = 1$ if desired - either way, $\perp$ will be included among the atomic formula. We call this extended language $\mathcal{L}$.

The system we use will be based on **G1c**, a well studied sequent system presented in chapter 3 of [12]. Amongst the permissible initial sequents we will include any sequents of the form:

(S1) $A \vdash A$

(S2) $\bot \vdash$

where A is an atomic formula of $\mathcal{L}$.[15]

The *structural rules* of the system will consist of left and right weakening, left and right contraction, and cut:

$$\frac{\Delta \vdash \Gamma}{A, \Delta \vdash \Gamma}\text{LW} \qquad \frac{\Delta \vdash \Gamma}{\Delta \vdash \Gamma, A}\text{RW}$$

$$\frac{A, A, \Delta \vdash \Gamma}{A, \Delta \vdash \Gamma}\text{LC} \qquad \frac{\Delta \vdash \Gamma, A, A}{\Delta \vdash \Gamma, A}\text{RC}$$

$$\frac{\Delta \vdash \Gamma, A \qquad A, \Delta' \vdash \Gamma'}{\Delta, \Delta' \vdash \Gamma, \Gamma'}\text{Cut}$$

Our systems inlcudes the *logical rules*:

$$\frac{\Delta, A \vdash \Gamma}{\Delta, A\&B \vdash \Gamma}\text{L\&} \qquad \frac{\Delta, B \vdash \Gamma}{\Delta, A\&B \vdash \Gamma}\text{L\&}$$

$$\frac{\Delta \vdash \Gamma, A \qquad \Delta \vdash \Gamma, B}{\Delta \vdash \Gamma, A\&B}\text{R\&}$$

$$\frac{\Delta, A \vdash \Gamma \qquad \Delta, B \vdash \Gamma}{\Delta, A \vee B \vdash \Gamma}\text{L}\vee$$

$$\frac{\Delta \vdash \Gamma, A}{\Delta \vdash \Gamma, A \vee B}\text{R}\vee \qquad \frac{\Delta \vdash \Gamma, B}{\Delta \vdash \Gamma, A \vee B}\text{R}\vee$$

$$\frac{\Delta \vdash \Gamma, A \qquad \Delta, B \vdash \Gamma}{\Delta, A \supset B \vdash \Gamma}\text{L}\supset \qquad \frac{\Delta, A \vdash \Gamma, B}{\Delta \vdash \Gamma, A \supset B}\text{R}\supset$$

$$\frac{\Delta, A[x/t] \vdash \Gamma}{\Delta, \forall x A \vdash \Gamma}\text{L}\forall \qquad \frac{\Delta \vdash \Gamma, A[x/a]}{\Delta \vdash \Gamma, \forall x A}\text{R}\forall$$

$$\frac{\Delta, A[x/a] \vdash \Gamma}{\Delta, \exists x A \vdash \Gamma}\text{L}\exists \qquad \frac{\Delta \vdash \Gamma, A[x/t]}{\Delta \vdash \Gamma, \exists x A}\text{R}\exists$$

where in the quantifier rules, x is an arbitrary variable, t is an arbitrary term (we count variables, constants, and anything composed from them using functions from the signature of the language as terms), and a is a variable - the so-called *eigenvariable* of the inference - and is assumed not to appear free in the bottom sequent in the rules $R\forall$ and $L\exists$. Here (as usual) $A[z/s]$, where z is a variable and s a term, is the result of replacing every free occurrence of z in A with s.

[15]See theorem 3.1.9 of [12] for a discussion of the restriction of (**S1**) to atomic formulae.

In each of the rules of inference, Δ and Λ are the *side formulae*. The remaining formula in the conclusion (i.e., the bottom sequent) is the *principal formula*, and the remaining formulae in the premises (i.e., the upper sequents) are the *active formulae* of the inference.[16]. This means, for example, that the rule of weakening has no active formula (though it has a principle formula), and the rule of cut no principal formula (though it has two instances of its active formula).

In this system, we regard the formula $\neg A$ as an abbreviation of $A \supset \bot$. The following rules may then be derived, and we use them freely:

$$\frac{\Delta \vdash \Gamma, A}{\Delta, \neg A \vdash \Gamma}\ \mathrm{L}\neg \qquad \frac{\Delta, A \vdash \Gamma}{\Delta \vdash \Gamma, \neg A}\ \mathrm{R}\neg$$

The introduction of the equality symbol requires the addition of the following initial sequents:

(S3$_s$) $\vdash s = s$

(S4$_{s_1,\ldots,t_n,f}$) $s_1 = t_1, ..., s_n = t_n \vdash f(s_1, ..., s_n) = f(t_1, ..., t_n)$

(S5$_{s_1,\ldots,t_n,R}$) $s_1 = t_1, ..., s_n = t_n, R(s_1, ..., s_n) \vdash R(t_1, ..., t_n)$

where s, the s_i, and the t_i are arbitrary terms, f any function and R any relation in the signature of the language $\mathcal{L}$.

Because the systems in which we are interested extend arithmetic, we also need the following initial sequents:

(S6) $x' = 0 \vdash$

(S7) $x' = y' \vdash x = y$

(S8) $\vdash x + 0 = x$

(S9) $\vdash (x + y)' = x + y'$

(S10) $\vdash x.0 = 0$

(S11) $\vdash x.y' = x.y + x$

Where here x and y are arbitrary variables. In saying that something is a permissible initial sequent, we also allow as an initial sequent any variant that may be obtained by renaming variables, or by substituting

[16]For example, in the rule of right & as just presented, the formula $A\&B$ in the right side of the lower sequent is the principal formula, and the formula A in the right side of the left upper sequent and the formula B in the right side of the right upper sequent are auxiliary formulae. The formulae in Δ and Γ are side formulae.

terms for free variables. So for example, given (**S6**), $y' = 0 \vdash$, $7' = 0 \vdash$ and $(x.y)' = 0 \vdash$ will all be permissible initial sequents.

Note that (**S8-11**) are just the defining equations for addition and multiplication. We also include amongst our mathematical initial sequents similar defining equations for all the primitive recursive functions we need (see §2.3 and Appendix 1 for a discussion of this.) As in axioms (**S8-11**), the formulae in these initial sequents may also be written in a purely equational form, so that they do not involve any connectives or quantifiers:

(S12) initial sequents giving the definitions of further primitive recursive functions.

We will also have a *rule* of mathematical induction I:

$$\frac{\phi[z/a], \Delta \vdash \phi[z/a'], \Gamma}{\phi[z/0], \Delta \vdash \phi[z/s], \Gamma}\, \mathrm{I}$$

where here the formula ϕ is an abritrary formula of $\mathcal{L}$ (and so may involve the truth predicate), x an arbitrary variable, s is an arbitrary term, and a is an arbitrary variable referred to as the *eigenvariable* of the inference. We require that a not appear free in $\phi[z/0], \Delta$ or Γ. In this rule, the formulae $\phi[z/a]$ and $\phi[z/a']$ are the active formulae, $\phi[z/0]$ and $\phi[z/s]$ are the principal formulae, and the formulae of Δ, Γ are the side formulae.

Now we turn to initial sequents that explicitly involve the truth predicate. In choosing these initial sequents, the idea is simply to capture the content of the axioms of $(\mathbf{MT}^{+})^{\circ}$. It will be important in what follows that these initial sequents contain only atomic formulae, and it is with this in mind that we choose our permitted initial sequents.

We begin with initial sequents that involve the truth predicate and the equality symbol. For all $n \geq 1$ we have:

(S13$_s$) $\vdash \mathbf{T}^{(n)}(\ulcorner s = s \urcorner)$

(S14$_{\vec{s},\vec{t}}$) $\vdash \mathbf{T}^{(n)}(\ulcorner \vec{s} = \vec{t} \supset f(\vec{s}) = f(\vec{t}) \urcorner)$

(S15$_{\vec{s},\vec{t}}$) $\vdash \mathbf{T}^{(n)}(\ulcorner \vec{s} = \vec{t} \supset [R(\vec{s}) \leftrightarrow R(\vec{t})] \urcorner)$

where s, the s_i and t_i are arbitrary terms. (These are axiom schemes.)

For each of the axioms (**S6-12**), there will be a corresponding axiom involving the truth predicate, where all the principal formulae are moved to the right side in the natural way. For all $n \geq 1$ we have:

(S16) $\vdash \mathbf{T}^{(n)}(\ulcorner x' \neq 0 \urcorner)$

(S17) $\vdash \mathbf{T}^{(n)}(\ulcorner x' \neq y' \supset x = y \urcorner)$

(S18) $\vdash \mathbf{T}^{(n)}(\ulcorner x + 0 = x \urcorner)$

(S19) initial sequents of the form $\vdash \mathbf{T}^{(n)}(\ulcorner ... \urcorner)$ corresponding to initial sequents (**S9-12**).

Again, with initial sequents (**S16-19**) and in the initial sequents that follow, we will assume that any variant that may be obtained by renaming variables or by substituting terms for free variables is also a permissible initial sequent.

We also have initial sequents concerning the truth of the principle of mathematical induction:

(S20$_\phi$) $\vdash \mathbf{T}^{(n)}(\ulcorner [\phi[z/0] \ \& \ (\forall x)(\phi[z/x] \supset \phi[z/x'])] \supset (\forall x)\phi[z/x] \urcorner)$

for any formula ϕ in the full language $\mathcal{L}$, where we assume that x does not appear free in $\phi[z/0]$.

We will also add some further mathematical initial sequents exclusively concerning the nature of truth predicate. In particular, we have the following for all $n \geq 1$:

(S21) $\mathbf{T}(0 = 1) \vdash$

(S22) $\vdash \mathbf{T}^{(n)}(\ulcorner \neg \mathbf{T}(0 = 1) \urcorner)$

(S23) $\text{Form}(x_1), \text{Form}(x_2), ... \vdash \mathbf{T}(s_i(x_1, x_2, ...))$

(S24) $\vdash \mathbf{T}^{(n)}(\ulcorner [\text{Form}(x_1) \ \& \ \text{Form}(x_2) \ \& ...] \supset \mathbf{T}(s_i(x_1, x_2, ...)) \urcorner)$

where in (**S23**) and (**S24**), each $s_i(x_1, x_2, ...)$ $(i = 1, ...k)$ is a primitive recursive function taking Gödel numbers of formulae to the Gödel number of a larger formulae corresponding to one of the logical axiom schemes of the Hilbert calculus, as spelt out in §2.3 and Appendix 1.

In addition, we have the following initial sequents for all $n \geq 1$:

(S25) $\text{Form}(x_1), \text{Form}(x_2), \mathbf{T}(x_1 \supset x_2), \mathbf{T}(x_1) \vdash \mathbf{T}(x_2)$

(S26) $\vdash \mathbf{T}^{(n)}(\ulcorner [\text{Form}(x_1) \ \& \ \text{Form}(x_2)] \supset [(\mathbf{T}(x_1 \supset x_2) \ \& \ \mathbf{T}(x_1)) \supset \mathbf{T}(x_2)] \urcorner)$

(S27) $\text{Form}(c), \text{Form}(a), \text{Var}(x), \text{Var}(z), \text{NotFree}(x, c), \mathbf{T}(c \supset a(x)) \vdash \mathbf{T}(c \supset (\forall z)a[x/z])$

(S28) $\vdash \mathbf{T}^{(n)}(\ulcorner [\text{Form}(c) \& \text{Form}(a) \& \ \text{Var}(x) \& \ \text{Var}(z) \& \ \text{NotFree}(x, c)] \supset [\mathbf{T}(c \supset a(x)) \supset \mathbf{T}(c \supset (\forall z)a[x/z])] \urcorner)$

(S29) $\text{Form}(c), \text{Form}(a), \text{Var}(x), \text{Var}(z), \text{NotFree}(x, c), \mathbf{T}(a(x) \supset c) \vdash \mathbf{T}((\exists z)a[z/x] \supset c)$

(S30) $\vdash \mathbf{T}^{(n)}(\ulcorner[\text{Form}(c)\&\text{Form}(a)\&\ \text{Var}(x)\&\ \text{Var}(z)\&\ \text{NotFree}(x,c)] \supset [\mathbf{T}(a(x) \supset c) \supset \mathbf{T}((\exists z)a[x/z] \supset c)]\urcorner)$

We call this system $\overline{\mathbf{MT}^{+}}$.

For a formula X in the language $\mathcal{L}$, we say that X is a *theorem* of $\overline{\mathbf{MT}^{+}}$ just in case there is a proof of the sequent $\vdash X$ in $\overline{\mathbf{MT}^{+}}$. We then have the following:

Lemma 3. *A formula X is a theorem of $\overline{\mathbf{MT}^{+}}$ iff it is a theorem of $(\mathbf{MT}^{+})^{\circ}$.*

Proof. First, we argue that if a sentence is a theorem of $\overline{\mathbf{MT}^{+}}$, then it is a theorem of $(\mathbf{MT}^{+})^{\circ}$.

Given a sequent μ of the form $\sigma_1, ..., \sigma_i \vdash \tau_1, ..., \tau_j$, let $S(\mu)$ be the sentence:

$$\neg(\sigma_1 \ \&...\&\ \sigma_i) \vee (\tau_1 \vee ... \vee \tau_j).$$

It then suffices to show that (i) for every initial sequent μ of $\overline{\mathbf{MT}^{+}}$, $S(\mu)$ is a theorem of $(\mathbf{MT}^{+})^{\circ}$, and (ii) if a rule of inference of $\overline{\mathbf{MT}^{+}}$ takes sequents $\mu_1, \mu_2, ...$ into ν, then if each $S(\mu_i)$ is a theorem of $(\mathbf{MT}^{+})^{\circ}$, then $S(\nu)$ is also a theorem of $(\mathbf{MT}^{+})^{\circ}$.

First we prove (i). This may be verified easily simply by going through the initial sequents (**S1-S30**) and checking that in each case the relevant sentence $S(\mu)$ is a theorem of $(\mathbf{MT}^{+})^{\circ}$. For example, (**S1**) and (**S2**) are logical truths, and so theorems of $(\mathbf{MT}^{+})^{\circ}$. Each of (**S3**) through (**S12**) is either identical with or a trivial logical consequence of **MT1** through **MT10** respectively, and thus theorems of $(\mathbf{MT}^{+})^{\circ}$. Likewise each of (**S13-S20**), is either identical with or a trivial logical consequence of $\mathbf{MT1}^{(n)}$ through $\mathbf{MT11}^{(n)}$ for $n > 1$, (**S21**) of **MT12**, (**S22**) of $\mathbf{MT12}^{(n)}$, (**S23**) of **MT13**, (**S24**) of $\mathbf{MT13}^{(n)}$, (**S25**) of **MT14**, (**S26**) of $\mathbf{MT14}^{(n)}$, (**S27**) of **MT15**, (**S28**) of $\mathbf{MT15}^{(n)}$, (**S29**) of **MT16**, and (**S30**) of $\mathbf{MT16}^{(n)}$, where $n > 1$.

The proof of (ii) is straightforward. One can easily very that with every structural and logical rule taking sequents $\mu_1, \mu_2, ...$ into ν, $s(\nu)$ is a logical consequence of $s(\mu_1), s(\mu_2), ...$, and thus if $s(\mu_1), s(\mu_2), ...$ are provable in $(\mathbf{MT}^{+})^{\circ}$, then so is $s(\nu)$. We consider finally the rule of induction. Suppose this rule takes the sequent $\phi[z/a], \Delta \vdash \phi[z/a'], \Gamma$ to $\phi[z/0], \Delta \vdash \phi[z/s], \Gamma$, where a does not occur free in Δ, Γ or $\phi[z/0]$. If we denote the upper sequent of this inference by μ and the lower sequent of this inference by ν, then the sentence $s(\mu)$ is logically equivalent to:

$$(\Delta\&\neg\Gamma) \supset (\phi[z/a] \supset \phi[z/a']).$$

If this is a theorem of $(\mathbf{MT}^+)^\circ$, and a does not occur free in Δ or Γ (as we are assuming), then in the Hilbert calculus presented in §2.3 we may immediately infer

$$(\Delta \& \neg\Gamma) \supset (\forall a)(\phi[z/a] \supset \phi[z/a']).$$

From this and (**MT11**), we may then infer:

$$(\Delta \& \neg\Gamma) \supset (\phi[z/0] \supset (\forall a)\phi[z/a])$$

where we use the fact that a does not appear free in $\phi[z/0]$. Using the fact that in the Hilbert calculus we have the logical axiom $(\forall a)\phi[z/a] \supset \phi[z/s]$, we then have:

$$(\Delta \& \neg\Gamma) \supset (\phi[z/0] \supset \phi[z/s])$$

which is logically equivalent to $s(\nu)$. Thus, if $s(\mu)$ is a theorem of $(\mathbf{MT}^+)^\circ$, so too is $s(\nu)$, as desired.

We now argue that if a sentence is a theorem of $(\mathbf{MT}^+)^\circ$, then it is a theorem of $\overline{\mathbf{MT}^+}$. It suffices to verify that for $n \geq 0$ and $i = 1, ..., 16$, each $\mathbf{MTi}^{(n)}$ is a theorem of $\overline{\mathbf{MT}^+}$. This is largely routine. The following derived rule is freely used in these arguments:

$$\dfrac{A_1, ... A_n \vdash B_1, ..., B_m}{\vdash (A_1 \& ... \& A_n) \supset (B_1 \vee ... \vee B_m)}\ \text{Der}$$

That this rule is indeed derivable is easily shown.[17].

Proofs of the sequent $\vdash \psi$ where ψ is $\mathbf{MTi}^{(0)}$ for $i = 1, ...10$ may be obtained either trivially or by applying this derived rule to the initial sequents (**S3-12**). If ψ is $\mathbf{MTi}^{(n)}$ for $i = 1, ...10$ and $n > 0$, $\vdash \psi$ may be proved from (**S13-19**). In similar ways, if ψ is $\mathbf{MTi}^{(n)}$ for $i = 12, ...16$ and $n = 0$ or $n \geq 0$, $\vdash \psi$ may be derived from the initial sequents (**S21-30**).

[17] For a proof, first note that we can derive the sequent $\Delta \vdash \Gamma, B_1 \vee B_2$ from $\Delta \vdash \Gamma, B_1, B_2$ as follows:

$$\dfrac{\dfrac{\dfrac{\Delta \vdash \Gamma, B_1, B_2}{\Delta \vdash \Gamma, B_1, B_2, B_1 \vee B_2}\ \text{RW} \qquad \dfrac{B_1 \vdash B_1}{B_1 \vdash B_1 \vee B_2}\ \text{R}\vee}{\Delta \vdash B_2, B_1 \vee B_2}\ \text{Cut, RW} \qquad \dfrac{B_2 \vdash B_2}{B_2 \vdash B_1 \vee B_2}\ \text{RW}}{\Delta \vdash B_1 \vee B_2}\ \text{Cut, RW}$$

Iterating this, it is easily shown that we can derive the sequent $\Delta \vdash B_1 \vee ... \vee B_m$ from $\Delta \vdash B_1, ..., B_m$. In a precisely analogous way, we can derive the sequent $A_1 \& ... \& A_n \vdash \Delta$ from $A_1, ..., A_n \vdash \Delta$. Combining these and using R$\supset$, one then gets a derivation of $\vdash (A_1 \& ... \& A_n) \supset (B_1 \vee ... \vee B_m)$ from $A_1, ... A_n \vdash B_1, ..., B_m$.

The only slightly non-trivial case to consider is if ψ is $\mathbf{MT11}^{(n)}$. If $n > 0$, a proof of $\vdash \psi$ is easily obtained from (**S20**). So the only case left to consider is if ψ is $\mathbf{MT11}^{(0)}$, i.e., ψ is:

$$(\phi[z/0]\ \&\ (\forall x)(\phi[z/x] \supset \phi[z/x'])) \supset (\forall x)\phi[z/x]$$

where x does not appear free in $\phi[z/0]$.

This sentence may be proved easily in $\overline{\mathbf{MT}^{+}}$:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\phi[z/x] \vdash \phi[z/x]}{\phi[z/x] \vdash \phi[z/x], \phi[z/x']}\,\text{RW} \qquad \dfrac{\phi[z/x'] \vdash \phi[z/x']}{\phi[z/x], \phi[z/x'] \vdash \phi[z/x']}\,\text{LW}}{\phi[z/x], \phi[z/x] \supset \phi[z/x'] \vdash \phi[z/x']}\,\text{L}\supset}{\phi[z/x], (\forall x)(\phi[z/x] \supset \phi[z/x']) \vdash \phi[z/x']}\,\text{L}\forall}{\phi[z/0], (\forall x)(\phi[z/x] \supset \phi[z/x']) \vdash \phi[z/x]}\,\text{I}}{\phi[z/0], (\forall x)(\phi[z/x] \supset \phi[z/x']) \vdash (\forall x)\phi[z/x]}\,\text{R}\forall$$

Even if ϕ is not atomic, the sequents $\phi[z/x] \vdash \phi[z/x]$ and $\phi[z/x'] \vdash \phi[z/x']$ may be easily derived. We may apply the rule I as shown, because x is not free in $(\forall x)(\phi[z/x] \supset \phi[z/x'])$ (as the variable x is bound in this expression), and nor is x free in $\phi[z/0]$ (by assumption.) For the same reasons, the subsequent R$\forall$ is also permissible. An application of the Der rule to the endsequent then gives a proof of $\mathbf{MT11}^{(0)}$.

This completes the proof that the theorems of $\overline{\mathbf{MT}^{+}}$ and $(\mathbf{MT}^{+})^{\circ}$ are the same. □

In light of Lemma 3, in order to prove Theorem 2′ it then suffices to prove:

Lemma 4. *For any closed sentence X, if $\overline{\mathbf{MT}^{+}} \vdash \mathbf{T}(\ulcorner X \urcorner)$, then $\overline{\mathbf{MT}^{+}} \vdash X$.*

In order to state and prove the main result we will use to prove this theorem, we will need some terminology. In the sequent calculus, the sequents in any individual inference can be divided into the *lowermost* sequent and *uppermost* sequent(s) in the obvious way. In an inference, we say that the principal formula(e) (which is always in the lowermost sequent) is an *immediate descendant* of any active formulae in the uppermost sequent(s), and that any side formula in the lowermost sequent is an immediate desecendant of the corresponding side formula(e) in the uppermost sequent.[18] [19] In a proof, say that an occurence of some formula θ' in some sequent is a *descendant* of an occurence of some formula θ in some (potentially different) sequent iff there is a sequence of occurences of formulae $\theta = \theta_0, \theta_1, ..., \theta_n = \theta'$ in which each θ_{i+1} is the immediate descendant of θ_i.

Say that a logical inference (i.e., an application of a logical rule) is *implicit* iff some descendant of its principlal formula is the cut formula in a cut inference. If a logical inference is not implicit, then some descendant of its principal formula can be found in the end-sequent, and we say that the logical inference is *explicit.*

The *endpiece* of a proof consists of all the sequents below which there is no implicit logical inference – i.e., it consists of all the sequents σ such that there is no implicit logical inference whose uppermost sequent co-incides with or lies properly below σ. (Thus, if a sequent is in the end piece, so is any sequent below it.) A cut is called *atomic* if its cut formula is atomic, and an application of weakening is called *atomic* if the formula it adds is atomic. (Note that in the system $\overline{\mathbf{MT}^+}$ all atomic formulae are of the form $\sigma = \tau$ or $T(\sigma)$, where σ and τ are (possibly open) terms.) Any inference whose lowermost sequent lies in the endpiece but whose uppermost sequents do not is a *boundary inference* of the proof.

Given any sequent $\Delta \vdash \Gamma$, for any variable x and term t we denote the sequent obtained from $\Delta \vdash \Gamma$ by replacing all free occurrences of x with t by $\Delta[x/t] \vdash \Gamma[x/t]$. Say the the set of permissible initial sequents of some system are *closed under term substitution* iff for any permissible initial sequent $\Delta \vdash \Gamma$, variable x and term t, $\Delta[x/t] \vdash \Gamma[x/t]$ is also a permissible initial sequent. (Note that we have assumed that the permissible initial sequents of $\overline{\mathbf{MT}^+}$ are closed under term substitution.) A simple induction on the construction of proofs shows that in any system in which the permissible initial sequents are closed under term substitution, whenever $\Delta \vdash \Gamma$ is provable, then for any variable x and term t, $\Delta[x/t] \vdash \Gamma[x/t]$ is also provable.[20]

All this terminology is standard in the proof theory literature, and is discussed in Chapter 2 of [11], for example.

[18] There may be more than such corresponding formula in the uppermost sequent in the case of the rule of contraction.

[19] Make sure principal / active formulae have been defined for structural rules, induction, and new truth rules.

[20] This argument is a variant of Lemma 3.5.2 of [12]. The only non-trivial step in the induction are the quantifier rules and induction. **Say something about these**

The main result we will use to prove Lemma 4 is:

Lemma 5. *Let V be a theory formulated in the sequent calculus with the following properties:*

(1). *The formulae in all initial sequents of V are atomic.*

(2). *The only rule of inference in V other than the structural and logical rules already outlined is that of mathematical induction, applied to some class of formulae.*

(3). *The permissible initial sequents of V are closed under term substitution.*

(4). *If σ is any closed term in the language and $\phi(x)$ any formula in the language $\mathcal{L}$ with a single free variable, then for some term n of the form $0''^{\cdots}$ (i.e., for some numeral n), there is a proof of the sequent $\phi(n) \vdash \phi(\sigma)$ in V in which every cut formula is atomic and there is no use of the rule of induction.*

Suppose that in V there a proof of a sequent in which every formula is atomic and closed (i.e., has no free variables.) Then there is a proof of the same sequent in which every sequent is atomic and closed, and the only rules used are cut, contraction and weakening.[21]

We begin by showing that Lemma 5 entails Lemma 4, and thus entails Theorems 2' and 2'.

Proof *of Lemma 4.* Note that the conditions of Lemma 5 are met by $\overline{\mathbf{MT}^+}$. (Condition (4) is the only non-trivial one, and may be proved by a straightforward induction.) We may thus apply the conclusion of Lemma 5 as needed.

For any formula s, let s^* be X if s has the form $\mathbf{T}(c)$, with c equal to the Godel number of a (possibly open) formula X, and $s^* = s$ otherwise. For a multiset of formulae M, let M^* just be the multiset consisting of the s^* for each s in M.

We prove that if a sequent $\Sigma \vdash \Gamma$ is provable in $\overline{\mathbf{MT}^+}$ using only cuts, weakenings and contractions, then $\Sigma^* \vdash \Gamma^*$ is also provable in $\overline{\mathbf{MT}^+}$. Because by the conclusion of Lemma 5 every sequent $\Sigma \vdash \Gamma$ consisting only of closed atomic formula that is provable in $\overline{\mathbf{MT}^+}$ is provable using only cuts, weakenings and contractions, it follows that if a sequent $\Sigma \vdash \Gamma$ consisting only of closed atomic formula is provable

[21]Note that the fact there is no application of the induction *rule* does not mean that there is no occurrence of the initial sequent (**S20**), if V allows it.

in $\overline{\mathbf{MT}^+}$, then $\Sigma^* \vdash \Gamma^*$ is also provable in $\overline{\mathbf{MT}^+}$. This in turn entails Lemma 4.

It suffices to prove (i) for each sequent $\Sigma \vdash \Gamma$ of the form (**S1-30**), $\Sigma^* \vdash \Gamma^*$ is provable in $\overline{\mathbf{MT}^+}$, and (ii) for each of the rules of cut, weakening and contraction in which the upper sequent(s) are $\Sigma_i \vdash \Gamma_i$ and lower sequent $\Sigma' \vdash \Gamma'$, if $\overline{\mathbf{MT}^+}$ proves each $\Sigma_i^* \vdash \Gamma_i^*$, then it also proves $\Sigma'^* \vdash \Gamma'^*$.

To prove (i), it suffices to go through each (**S1-30**) case by case. The case of (**S1**) is trivial. With the exception of (**S5**), initial sequents (**S2-12**) do not involve the truth predicate, and thus for each of these initial sequents $\Sigma \vdash \Gamma$, we have $\Sigma^* = \Sigma$ and $\Gamma^* = \Gamma$. It thus follows immediately that $\Sigma^* \vdash \Gamma^*$ is provable in $\overline{\mathbf{MT}^+}$. In the case of (**S5**) in which R is the truth predicate, it suffices to prove that if t_1, t_2 are closed terms with t_1 equal to the Gödel number of a sentence X_1, and t_2 equal to the Gödel number of a sentence X_2, then the sequent $t_1 = t_2, X_1 \vdash X_2$ is provable in $\overline{\mathbf{MT}^+}$. But if t_1 and t_2 have the same numeric value, then $X_1 = X_2$, and the sequent $t_1 = t_2, X_1 \vdash X_2$ follows from an instance of (**S1**) and weakening, while if t_1 and t_2 have different numeric values, then $\overline{\mathbf{MT}^+}$ proves $t_1 = t_2 \vdash$, and so $t_1 = t_2, X_1 \vdash X_2$ follows by weakening.

If $\Sigma \vdash \Gamma$ is one of the initial sequents (**S13-20**), then if $n > 1$, $\Sigma^* \vdash \Gamma^*$ is also of the form (**S13-20**), and thus provable in $\overline{\mathbf{MT}^+}$. In the case $n = 1$, $\Sigma^* \vdash \Gamma^*$ has the form of one of the intial sequents (**S3-S12**) or follows easily from such a sequent (using a rule such as Der), and is thus also provable in $\overline{\mathbf{MT}^+}$.

In the case of (**S21**), it suffices to show that $\overline{\mathbf{MT}^+}$ proves $0 = 1 \vdash$, which is straightforward. In the case of (**S22**) with $n > 1$ the result is trivial, and in the case $n = 1$ it suffices to show that $\overline{\mathbf{MT}^+}$ proves $\vdash \neg\mathbf{T}(0 = 1)$, which follows from (**S21**). In the case of (**S23**), it suffices to to show that if $c_1, c_2, ...$ are the Gödel numbers of formulae in $\mathcal{L}$, then $\overline{\mathbf{MT}^+}$ proves $\mathrm{Form}(c_1), \mathrm{Form}(c_2), ... \vdash S_i$, where S_i is an instance of the relevant logical axiom of the Hilbert calculus. But S_i is a tautology and so $\overline{\mathbf{MT}^+}$ proves $\vdash S_i$, and thus by weakening, also proves $\mathrm{Form}(c_1), \mathrm{Form}(c_2), ... \vdash S_i$. The case of (**S24**) is dealt with similarly.

In the case of (**S25**) it suffices to show that if c_1, c_2 are the Gödel numbers of formulae X_1, X_2 in $\mathcal{L}$, then $\overline{\mathbf{MT}^+}$ proves $\mathrm{Form}(c_1), \mathrm{Form}(c_2), X_1 \supset X_2, X_1 \vdash X_2$. This is immediate from the fact that $\overline{\mathbf{MT}^+}$ proves $X_1 \supset X_2, X_1 \vdash X_2$ for each pair of sentences X_1, X_2. The case of (**S26**) is dealt with similarly.

In the case of (**S27**), it suffices to show that if c, a are the Gödel number of formulae C, A, and x and z the Gödel number of variables X and Z, then $\overline{\mathbf{MT}^{+}}$ proves the sequent:

$$\mathrm{Form}(c), \mathrm{Form}(a), \mathrm{Var}(x), \mathrm{Var}(z), \mathrm{NotFree}(x, c), C \supset A(X) \vdash C \supset (\forall Z)A[X/Z]. \qquad (\ddagger)$$

But if x does not occur free in c, then $C \supset A(X) \vdash C \supset (\forall Z)A[X/Z]$ is provable in $\overline{\mathbf{MT}^{+}}$, while if x does not occur free in c, then $\mathrm{NotFree}(x, c) \vdash$ is provable in $\overline{\mathbf{MT}^{+}}$, and so either way, ($\ddagger$) may be obtained by weakening. The cases of (**S28-30**) may all be dealt with similarly.

We now prove (ii). These arguments are all trivial. For example, consider the following instance of cut:

$$\frac{\Delta \vdash \Gamma, A \qquad A, \Delta' \vdash \Gamma'}{\Delta, \Delta' \vdash \Gamma, \Gamma'}\,\mathrm{Cut}$$

Then the following inference is also a legitimate application of cut:

$$\frac{\Delta^{*} \vdash \Gamma^{*}, A^{*} \qquad A^{*}, \Delta'^{*} \vdash \Gamma'^{*}}{\Delta^{*}, \Delta'^{*} \vdash \Gamma^{*}, \Gamma'^{*}}\,\mathrm{Cut}$$

and thus, if the upper sequents are provable in $\overline{\mathbf{MT}^{+}}$, then so is the lower sequent. Equally trivially considerations apply in the case of contraction and weakening. □

At this point, it suffices to prove Lemma 5. The proof is a straightforward repurposing of Gentzen's argument for the consistency of arithmetic (see [6], [7].) We will mimic this argument as presented by Takeuti [11]. Much of the argument here is identical to the argument found there, and so in the interest of brevity we will at several points simply summarize the argument and refer the reader there.

Proof of Lemma 5. The argument revolves around an assignment of ordinals to proofs in $\overline{\mathbf{MT}^{+}}$. To define this assignment, we first need several definitions. The *grade* of a formula is the number of logical symbols ($\vee$, &, $\supset$, $\forall$, $\exists$) that occur in it. The *grade of a cut* (resp. *induction*) is the grade of the associated cut formula (resp. grade of a forrmula on which the induction is performed.)[22] The *height* of

[22]By induction here we mean exclusively the application of the *rule* of induction, and not the use of an initial sequent like (**S20**).

a sequent is the maximum grade of any cut or induction appearing below the sequent in question.

To each sequent μ in a proof an ordinal $o(\mu)$ is then assigned as follows. If μ is an initial sequent, then $o(\mu)$ is 0. In any weakening or contraction, if μ is the upper sequent of the inference and ν the lower sequent, then $o(\nu) = o(\mu)$. In any unary logical inference (i.e., L&, R$\vee$, R$\supset$, L$\forall$, R$\forall$, L$\exists$, R$\exists$), if μ is the upper sequent of the inference and ν the lower sequent, then $o(\nu) = o(\mu) + 1$. In any binary logical inference (i.e., R&, L$\vee$, L$\supset$), if μ_1, μ_2 are the upper sequents of the inference and ν the lower sequent, then $o(\nu) = o(\mu_1)\#o(\mu_2)$.

In any cut, if μ_1, μ_2 are the upper sequents and ν the lower sequent, then let k be the height of the sequents μ_1, μ_2 (these sequents will have the same height), and l be the height of the lower sequent. We then have $k \geq l$. We define:

$$o(\nu) = \omega_{k-l}(o(\mu_1)\#o(\mu_2))$$

where $\omega_0(\alpha) = \alpha$ and $\omega_{i+1}(\alpha) = \omega^{\omega_i(\alpha)}$.

In the case of induction, we first need some preliminary notation. For any ordinal $\alpha \leq \epsilon_0$, α can be written uniquely in the normal form $\omega^{\alpha_1} + \omega^{\alpha_2} + ... + \omega^{\alpha_n}$ where $\alpha_1 \geq \alpha_2 \geq ... \geq \alpha_n$. For any α, α_1 thus denotes the largest exponent in this sum.

In any induction, if μ is the upper sequent and ν the lower sequent, and k and l are their respective heights, we then define

$$o(\nu) = \omega_{k-l+1}(o(\mu)_1 + 1).$$

Iterating these definitions over the construction of a proof, we then define the ordinal of a proof to be that its final sequent. (We here use condition (2) - i.e., the fact that the only non-structural, non-logical rule is that of mathematical induction.)

Suppose that a sequent μ is provable in V, where every formula in μ is closed and atomic. Of all the proofs of μ, pick the one π whose ordinal is least. We will show that there is a proof of μ with the same ordinal, in which every sequent is atomic and closed, and the only rules used are cut, contraction and weakening

Because every formula in μ is atomic, it follows that no logical rule can occur in the endpiece of π. (This is because only atomic formulae appear in the endsequent, and so no logical inference in π can be explicit - i.e., all logical inferences in π are implicit.) This means that the endpieceof π just consists of all sequents below which there is no logical inference, and so the endpiece of π thus consists at most of applications of weakening, contraction, cut and induction.

We now show that all formulae in the endpiece of π may be taken to be closed. This is true of the endsequent by assumption. For the rules of weakening and contraction, if all the formulae in the lower sequent are closed, then all the formulae in the upper sequent are also closed. Take a sequent S in the endpiece with an open variable, such that no open variable occurs in the proof below S. We know S is not the endsequent, so S is the upper sequent of some cut or induction.

Suppose S is the upper sequent of a cut. This means that the cut formula is open, but all other formulae in the inference are closed. Suppose the cut formula is $\theta(x)$ with free variable x. The cut has the form:

$$\frac{\Delta_1 \vdash \Gamma_1, \theta(x) \qquad \theta(x), \Delta_2 \vdash \Gamma_2}{\Delta_1, \Delta_2 \vdash \Gamma_1, \Gamma_2}$$

where all the formulae of the Δ_i, Γ_i are closed. We now use the fact that whenever a sequent $\Delta \vdash \Gamma$ is provable, then for any term t so is the sequent $\Delta[x/t] \vdash \Gamma[x/t]$ (i.e., the sequent in which all free occurences of x are replaced by the term t) without increasing the ordinal of the proof. (This is a straightforward proof by induction on the construction on proofs. The base case is established by using condition (2) - the fact that the permissible intiial sequents of V are closed under term substitution.[23].) We use this fact to remove the free variable x in the cut as follows, without altering the lower sequent or anything below it:

$$\frac{\Delta_1 \vdash \Gamma_1, \theta(0) \qquad \theta(0), \Delta_2 \vdash \Gamma_2}{\Delta_1, \Delta_2 \vdash \Gamma_1, \Gamma_2}$$

(where here we have let t be the constant term 0.)

Suppose then that S is an application of the rule of induction,

$$\frac{F(a), \Delta \vdash F(a'), \Gamma}{F(0), \Delta \vdash F(s), \Gamma}\,\mathrm{I}$$

here the upper sequent contains the free variable a,

is it also possible for the upper sequent to contain a free variable while the lower sequent contains no free variable - in particular, the upper sequent always contains the free variable a, while if s is a closed term, and Δ, Γ consist only of closed formulae, then the lower sequent

[23]This argument is a variant of Lemma 3.5.2 of [12]

contains no free variable. Suppose then in some application of induction, there is no free variable in the lower sequent. Because s is a closed term, by condition (4) in the Lemma, it follows that for some numeral m (i.e., term of the form $0''..'$), there is a proof of $F(m) \vdash F(s)$ without non-atomic cuts or inductions.

Here we use again the fact that whenever a sequent $\Delta \vdash \Gamma$ is provable, so too is the sequent $\Delta[x/t] \vdash \Gamma[x/t]$. Using this, it follows that whenever $F(a), \Delta \vdash F(a'), \Gamma$ is provable, so too is the sequent $F(n), \Delta \vdash F(n'), \Gamma$ for each numeral n. From this fact it is then possible to prove the sequent $F(0), \Delta \vdash F(s), \Gamma$ without induction using a sequence of cuts. For example, in the case $m = 0'''$, we have:

$$\dfrac{\dfrac{\dfrac{\vdots}{F(0), \Delta \vdash F(0'), \Gamma} \quad \dfrac{\vdots}{F(0'), \Delta \vdash F(0''), \Gamma}}{F(0), \Delta \vdash F(0''), \Gamma}\text{Cut} \quad \dfrac{\vdots}{F(0''), \Delta \vdash F(0'''), \Gamma}}{F(0), \Delta \vdash F(0'''), \Gamma}\text{Cut}$$

where for brevity we have omitted various applications of contraction. From the fact that there is a proof of $F(m) \vdash F(s)$ we may then append the following to the above proof (using again $m = 0'''$):

$$\dfrac{F(0), \Delta \vdash F(0'''), \Gamma \qquad F(0'''), \Delta \vdash F(s), \Gamma}{F(0), \Delta \vdash F(s), \Gamma}\text{c}$$

Eliminating the instance of induction in question, we obtain a proof π' with the same endsequent as π. It may be shown that the ordinal of π' is properly smaller than of π (see pp. 105-6 of [11].) It follows then that if π is indeed the proof of the given endsequent with the lowest ordinal, then π can have no induction in its endpiece. This also completes the argument that we may assume that all formulae in the endpiece of π are closed.

To summarize, we have shown that we may assume the endpiece of π consists at most of applications of weakening, contraction and cut, and contains only closed formulae.

Next, we show that any weakening in the endpiece must be atomic. Suppose there is a non-atomic weakening in the endpiece, and that σ is the non-atomic formula introduced in this weakening. Because the only rules of inference in the endpiece are weakening, contraction and cut, and because σ does not appear in the end-sequent (being non-atomic), it must be the cut formula in some cut. We may assume without loss of generality that the cut in question appears immediately after the weakening in question, i.e., the endpiece contains something like the following steps:

$$\dfrac{\dfrac{\Delta \vdash \Gamma}{\Delta, A \vdash \Gamma}\text{ LW} \qquad \Delta' \vdash \Gamma', A}{\Delta, \Delta' \vdash \Gamma, \Gamma'}\text{ Cut}$$

But this cut may be removed, and replaced by a series of weakenings:

$$\dfrac{\Delta \vdash \Gamma}{\Delta, \Delta' \vdash \Gamma, \Gamma'}$$

One can easily argue that the ordinal of the resulting proof is strictly lower than the ordinal of the original proof. This contradicts our assumption that π had the lowest ordinal of any proof of the endsequent. Thus, there can be no non-atomic weakening in the endpiece of π.

We have shown that we may assume the endpiece of π consists at most of applications of atomic weakening, contraction and cut, and contains only closed formulae. We show now that all cuts in the endpiece are atomic.

To do so, we first define a cut in the endpiece of a proof to be *suitable* iff both occurrences of the cut formula in the proof are descendants of the principal formula of (different) boundary inferences. One may then prove the lemma that for any proof ρ, if (i) ρ is not its own endpiece, (ii) the endpiece contains only contractions, cuts, and atomic weakenings, and (iii) any initial sequent used in the proof contains only atomic formulae, then there exists a suitable cut in the end-piece of ρ. This is just (a minor variation of) sublemma 12.9 of [11], and we refer the reader there for the details of the proof.[24]

---

[24] In summary, the argument is by induction on the number of non-atomic cuts in the endpiece of ρ. If there are no non-atomic cuts in the endpiece, then there can be no boundary inferences. Thus, ρ is its own endpiece, and so the theorem vacuously obtains. Suppose the theorem has been proved whenever the endpiece has $< n$ non-atomic cuts, and suppose ρ has n non-atomic cuts. Take the lowest such non-atomic cut:

$$\dfrac{\Delta \vdash \Gamma, A \qquad A, \Delta' \vdash \Gamma'}{\Delta, \Delta' \vdash \Gamma, \Gamma'}\text{ Cut}$$

where here A is non-atomic. Assume this cut in not essential; this means that the A in one of the uppermost sequents of the inference shown is not descended from the principal formula of a boundary inference. Without loss of generality, suppose it is the occurrence of A in the left uppermost sequent of this cut that has this property. Because there are no non-atomic weakenings in the endpiece of ρ, this occurence of A must either be a descendant of a formula in an initial sequent, or a descendant of a non-principal formula in a boundary inference. The former is impossible, because all formulae in initial sequents

Suppose then that there is a non-atomic cut in the endpiece of π. Then condition (i) of the theorem just cited holds, as the cut formula must be a descendant of some formulae in a boundary inference, as all formulae in all initial sequents are atomic. (If the endpiece has a boundary inference, then the endpiece must be a proper subset of the original proof.) Condition (ii) holds as a result of the argument given so far, and condition (iii) follows from condition (1) of the lemma. Thus the endpiece of π contains a suitable cut. But suitable cuts may be eliminated by standard techniques, as explained in detail in Chapter 2 of [11] or throughout [12]. Suppose, for example, that the cut formula in a suitable cut is a disjunction $A \vee B$. Then π has something like the following form:

$$\dfrac{\dfrac{\dfrac{\dfrac{\Delta_1 \vdash \Gamma_1, A}{\Delta_1 \vdash \Gamma_1, A \vee B}\,\mathrm{R}\vee}{\vdots}}{\Delta \vdash \Gamma, A \vee B} \qquad \dfrac{\dfrac{\dfrac{A, \Delta_2 \vdash \Gamma_2 \qquad B, \Delta_2 \vdash \Gamma_2}{A \vee B, \Delta_2 \vdash \Gamma_2}\,\mathrm{L}\vee}{\vdots}}{A \vee B, \Delta' \vdash \Gamma'}}{\Delta, \Delta' \vdash \Gamma, \Gamma'}\,\mathrm{Cut}$$

where the uppermost inferences in this proof are boundary inferences of the endpiece of π, and other than the inferences shown, every other inference is a contraction, atomic weakening, or cut. This may be replaced by:

$$\dfrac{\Delta_1 \vdash \Gamma_1, A \qquad A, \Delta_2 \vdash \Gamma_2}{\Delta_1, \Delta_2 \vdash \Gamma_1, \Gamma_2}\,\mathrm{Cut}$$

and precisely the same sequence of contractions, atomic weakenings, and cuts may be used to derive $\Delta, \Delta' \vdash \Gamma, \Gamma'$ from this. The ordinal of this new proof may be shown to be strictly less than that of π (see, for example Chapter 2, §12 of [11].) Standard arguments show that this holds regarless of the form of the non-atomic cut we are assuming to exist in the endpiece of π. This contradicts our assumption that the

are atomic. So this occurence of A must be a descendant of a non-principal formula in a boundary inference. Consider then subproof ρ' of ρ that ends in $\Delta \vdash \Gamma, A$. It is easily shown that the endpiece of ρ' is just the intersection of the the endpiece of ρ and ρ'. (This relies on the fact that there is no non-atomic cut below the cut we have been considering.) Thus, because it contains a boundary inference, condition (i) of the theorem obtains for ρ', as do conditions (ii) and (iii). Because the endpiece of ρ' has $< n$ non-atomic cuts, it follows by inductive hypothesis that ρ' contains a suitable cut. This of course is also a suitable cut of ρ.

ordinal of π is minimal. It thus follows that there is no non-atomic cut in pi.

We have therefore shown that the endpiece of π consists at most of applications of atomic cuts, contractions, and atomic weakenings, and contains only closed formulae. Because all cuts are atomic and the endsequent contains only atomic formulae, it follows that there are no boundary inferences, and so the endpiece of π is π in its entirety. Thus π in its entirity consists only of atomic cuts, contractions, and atomic weakenings and contains only closed formulae. Clearly π can contain no non-atomic formulae, as such a formula could only be eliminated by a non-atomic cut. Thus in π every sequent is atomic and closed, and the only rules used are cut, contraction and weakening. This completes the proof of Lemma 5. □

References

[1] Buss, S., [1998], 'An introduction to proof theory', Chapter 1 of *Handbook of Proof Theory*, Elsevier.

[2] Feferman, S., [1962], Transfinite Recursive Progressions of Axiomatic Theories , The Journal of Symbolic Logic

[3] Field, H., [2006], Truth and the Unprovability of Consistency, Mind, Vol. 115, p. 459.

[4] Franzen, T. [2004], *Inexhaustibility*, Lecture Notes in Logic, No. 16, Association for Symbolic Logic.

[5] Friedman, H. and Sheard, M., [1987] "An Axiomatic Approach to Self-Referential Truth", Annals of Pure and Applied Logic, 33: 1-21.

[6] Gentzen, G., [1936], Die Widerspruchsfreiheit der reinen Zahlentheorie, Mathematische Annalen, 112: 493?565, tranislated as "The consistency of arithmetic", in Gentzen, Szabo (1969).

[7] Gentzen, G., Szabo, M. (ed.) [1969], Collected Papers of Gerhard Gentzen, Studies in logic and the foundations of mathematics , Amsterdam.

[8] Halbach, V., [2011], "Axiomatic Theories of Truth", Cambridge University.

[9] Heck, "Truth and Disquotation", Synthese, Vol. 142, No. 3, Tarski on Truth: A Centennial Celebration (Dec., 2004), pp. 317 -352.

[10] Negri, S. and von Plato, J. [2001] Structural Proof Theory, Cambridge University Press.

[11] Takeuti, G., [1987] Proof Theory, Second Edition, North-Holland.

[12] Troelstra, A. and Schwitchtenberg, H. [2000] Basic Proof Theory, Second Edition, Cambridge University Press.