The Journal of Symbolic Logic
Volume 89, Number 1, March 2024

CLASSICAL DETERMINATE TRUTH I

KENTARO FUJIMOTO AND VOLKER HALBACH

Abstract. We introduce and analyze a new axiomatic theory CD of truth. The primitive truth predicate can be applied to sentences containing the truth predicate. The theory is thoroughly classical in the sense that CD is not only formulated in classical logic, but that the axiomatized notion of truth itself is classical: The truth predicate commutes with all quantifiers and connectives, and thus the theory proves that there are no truth value gaps or gluts. To avoid inconsistency, the instances of the T-schema are restricted to *determinate* sentences. Determinateness is introduced as a further primitive predicate and axiomatized. The semantics and proof theory of CD are analyzed.

> Ad veritatem autem copulativae requiritur quod utraque pars sit vera, et ideo si quaecumque pars copulativae sit falsa, ipsa copulativa est falsa.
>
> William of Ockham, *Summa Logicae* II.32

> Ad veritatem autem [propositionis] disiunctivae requiritur quod aliqua pars sit vera[.]
>
> William of Ockham, *Summa Logicae* II.33

§1. Classicality and compositional semantics. Philosophy abounds with general claims expressed with a truth predicate: Philosophers debate whether there are contingent or synthetic a priori truths, and whether there are unprovable or unverifiable truths; they mostly agree that what is known is true, but that some justified true beliefs are not known; they teach their students that the conclusion of a valid argument is true if its premisses are true and try to convince their students that the rules of natural deduction are truth-preserving. Specific instances of these claims can often be stated without a predicate for truth; but the quantified claims make essential use of the truth predicate.

In the arguments for this kind of claim, assumptions about the truth predicate are often used implicitly without further ado. In the present paper we strive to make these assumptions explicit by listing principles about the truth predicate that are jointly consistent with further plausible assumptions about the objects to which truth is ascribed. We presuppose that these objects share the structure of (types of)

Received November 5, 2021.
2020 *Mathematics Subject Classification.* Primary 03F03.
Key words and Phrases. axiomatic theories of truth, ordinal analysis, liar paradox, determinateness, grounding.

0022-4812/24/8901-0013
DOI:10.1017/jsl.2023.49

sentences without entering the debate about whether truth applies to propositions, beliefs, sentence tokens, or still other things.[1]

It is widely believed that truth serves its role as a device of generalization in virtue of its disquotational feature; and a large part of the literature on truth is devoted to the development of strong disquotational truth theories. However, in their zeal to devise theories of disquotational truth, some philosophers seem to have lost sight of how we reason with generalizations. When truth is used as a device of generalization, disquotation is not the only principle that is used in reasoning with generalizations. Compositional principles are just as important and used routinely without qualms. We give three examples.

Our first example comes from logic. When motivating axioms or rules for a logical calculus, compositional laws of truth are used. In the case of the elimination rule for conjunction or for the universal quantifier in natural deduction, students usually quickly accept the claim that the rule preserves truth. At this point in the course, students may not have seen the model-theoretic notions of satisfaction and truth or the object/metalanguage distinction at all; nevertheless they have no problems in seeing that a conjunct must be true if the conjunction is true and that a substitution instance is true if the universally quantified sentence is true. The point can also be made about the history of logic: Logicians chose sound logical calculi long before the arrival of the modern model-theoretic notion of truth in Tarski and Vaught [43] in 1956 (as far as we know; see Hodges [26]). Truth preservation of logical rules is often assumed to hold without any restriction. Any restriction to a specific sublanguage is at odds with the universality of logic.

The use of the compositional principles is by no means confined to the realm of logical theorizing. Our second example is from epistemology. Most epistemologists would be happy to endorse the following argument, which could form part of a Gettier example:

> Smith believes a disjunction. He is justified in believing one disjunct, which happens to be false, while he does not believe the other disjunct, which happens to be true. Therefore Smith has a justified true belief.

In this argument we reason about the truth (and justification) of beliefs by analyzing and manipulating their logical structures without explicitly specifying exactly what these beliefs are. The validity of the argument depends on the assumption that a disjunction is true if one of its disjuncts is true, as stated by Ockham in the second quote above. Fujimoto [16, 17] called this kind of reasoning *blind deduction*, because we reason about the truth of beliefs or sentences without being able to specify these beliefs or sentences by means of a quotational name or a structural description. Disquotation axioms or T-sentences used as axioms permit only reasoning about the truth of sentences that are explicitly given by such a naming device.

[1] There are many alternative approaches. Philosophers have tried to recover the expressive power of the truth predicate by employing propositional quantification. Sophisticated theories of propositions as objects to which truth is ascribed have been developed. Here, we do not have space to discuss these alternatives. See Halbach and Leigh [24] for a discussion.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

The use of the compositional principles is so ubiquitous that their use often goes unnoticed as in our third example:

> Alfred made a claim that was denied by Kurt (i.e., Kurt asserted the negation of the claim); and everything Kurt claimed is true. Therefore not everything Alfred claimed is true.

For the validity of this argument we require the principle that a claim is not true if its negation is true.

In the examples we quantify over *all* sentences, beliefs, or propositions. Blind deductions are not restricted to sentences that are safe, grounded, determinate, healthy, non-circular, or in some other way 'unproblematic'. Moreover, the full compositional principles are consistent without any restrictions on the underlying logic or syntax theory.

The axioms for truth can and should be added to a base theory that yields at least a comprehensive theory of syntax (in a direct or coded form); the base theory may also go far beyond a theory of syntax. In [22, 23] the second author formulated the truth theory over set theory as the base. Here, however, we start from Peano arithmetic, which is traditionally used as base theory for axiomatic theories of truth; but we consider it only as a simple model case. Expressions are identified with their codes. As usual, we confine ourselves to a truth rather than a satisfaction predicate in the case of arithmetic, as names for all objects are available and satisfaction and variable assignments are not needed. All extensions of Peano arithmetic we consider are formulated in classical logic.

In what follows, $\mathcal{L}_0$ is the language of arithmetic. The axioms are formulated in an expansion of $\mathcal{L}_0$ with further predicate symbols T and D. Negation and conjunction are the only connectives, and the universal is the only quantifier in the language; other connectives and the existential quantifier are assumed to be defined. In the axioms below $\mathrm{Sent}(x)$ expresses that x is a sentence of $\mathcal{L}$. The symbol $\dot{\neg}$ expresses the function that, if it is applied to a sentence, returns its negation. If a suitable function symbol is not in $\mathcal{L}_0$, this function needs to be expressed by a suitable formula; $\dot{\wedge}$ and $\dot{\forall}$ are defined analogously. The quantifier $\forall t$ ranges over (codes of) closed terms and is defined using a suitable formula describing the set of closed terms; and $x(t/v)$ designates the result of formally substituting t for the variable v in x. The following axioms then express that a negated sentences is true iff the sentence is not, that a conjunct is true iff both conjuncts are, and that a universally quantified sentence is true iff all its substitution instances are:

$\mathsf{T4}\ \forall x \left(\mathrm{Sent}(x) \rightarrow (\mathrm{T}(\dot{\neg}x) \leftrightarrow \neg \mathrm{T}x)\right)$,
$\mathsf{T5}\ \forall x \forall y \left(\mathrm{Sent}(x \dot{\wedge} y) \rightarrow (\mathrm{T}(x \dot{\wedge} y) \leftrightarrow \mathrm{T}x \wedge \mathrm{T}y)\right)$,
$\mathsf{T6}\ \forall v \forall x \left(\mathrm{Sent}(\dot{\forall} v\, x) \rightarrow (\mathrm{T}(\dot{\forall} v\, x) \leftrightarrow \forall t\, \mathrm{T}x(t/v))\right)$.

The axioms T4–T6 capture a *thoroughly* classical concept or truth: The truth theory is not only formulated in classical logic, but the notion of truth axiomatized by T4–T6 is itself classical. According to the axioms, truth commutes with quantifiers and connectives for *all* sentences, including those with the truth predicate. Theories such as the Kripke–Feferman theory [8, 37] or Cantini's [6] VF are formulated in classical logic as well, but capture a non-classical notion of truth.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

Axiom T4 disproves the existence of truth value gaps and gluts. We define falsity as the truth of the negation, that is, we stipulate: $\mathrm{F}x :\Leftrightarrow \mathrm{T}\dot{\neg}x$. If the truth predicate is classical and T4 is assumed, then falsity is equivalent to non-truth. Thus, every sentence is true or false; and no sentence is both true and false.

Disquotation axioms have been been thought to be more fundamental than the compositional axioms. Consequently, there have been various attempts to obtain compositional from disquotational principles. Tarski [42, p. 259] tried to derive restricted compositional principles from the T-sentences using what he called the *rule of infinite induction*, which is a generalization of the ω-rule. More recently, some authors, including Halbach [19] and Horsten and Leigh [27], have used reflection principles instead, which are of course finitized versions of infinitary rules. We are sceptical about the prospects of justifying compositional from disquotational axioms. There are general objections against the use of infinitary and reflection rules to this end. Another crucial objection is especially relevant when unrestricted compositional principles are employed. Usually, attempts to obtain compositional axioms for *all* sentences of a certain kind from disquotational principles rely on the disquotation principle for all sentences of this kind. For instance, we may try to derive the compositional axioms for all T-free (or T-positive) sentences from all equivalences $\mathrm{T}\ulcorner\phi\urcorner \leftrightarrow \phi$, where ϕ is T-free (or T-positive). However, in the case of the *completely unrestricted* compositional axioms T4–T6, this strategy is not very promising. Of course, the unrestricted compositional axioms T4–T6 can be derived from all instances $\mathrm{T}\ulcorner\phi\urcorner \leftrightarrow \phi$; but this is because the unrestricted T-schema is inconsistent over arithmetic. Hence, we will not be able to derive all type-free compositional from disquotational principles, even if some kind of infinitary rule is assumed.[2] Therefore we adopt the unproblematic, fully general compositional principles as axioms and do not attempt to derive them from a disquotation schema, which has to be restricted in some way.

§2. Determinateness and disquotation. Our policy for restricting disquotation is inspired by disquotationalist and deflationist theories of truth. We expand the base language $\mathcal{L}_0$ by adding sentences that are needed for semantic ascent and generalization. The sentences of the base language together with those needed for semantic ascent and generalization form the set $\mathfrak{D}$. Roughly, if we have a sentence ϕ in $\mathfrak{D}$, we also add an equivalent new sentence $\mathrm{T}\ulcorner\phi\urcorner$ as a 'copy' of ϕ. These copies are required for semantic ascent. We close then under connectives and quantifiers in a way to be explained. This permits generalizations, because we can now express, for instance, that all provable sentences of $\mathcal{L}_0$ are true. This process is then iterated: We add copies $\mathrm{T}\ulcorner\phi\urcorner$ for semantic ascent and then close under connectives and quantifiers in order to be able to generalize. We add only sentences to $\mathfrak{D}$ that

[2]This claim requires some qualification. It is possible to obtain all compositional axioms from a single consistent instance of the T-schema by a trick due to McGee [32] based on Curry's paradox. Let C be the conjunction of all compositional axioms. By the diagonal lemma there is a sentence γ such that $\gamma \leftrightarrow (\mathrm{T}\ulcorner\gamma\urcorner \leftrightarrow C)$ is provable. This sentence is logically equivalent to $(\mathrm{T}\ulcorner\gamma\urcorner \leftrightarrow \gamma) \leftrightarrow C$. Therefore the equivalence $\mathrm{T}\ulcorner\gamma\urcorner \leftrightarrow \gamma$ is a re-axiomatization of the compositional axioms over arithmetic. We see this as a mere curiosity; but it shows that the claim that one cannot get all type-free compositional axioms from a consistent set of T-sentences needs some qualification.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

are required for semantic ascent and generalization, not all sentences of the full language $\mathcal{L}$. In particular, not all sentences of the form $\mathrm{T}\ulcorner\phi\urcorner$ are in $\mathfrak{D}$. The set $\mathfrak{D}$ fails to be decidable. In a general setting with 'contingent' vocabulary, whether a sentence belongs to $\mathfrak{D}$ may depend on empirical facts.

Our truth theory CD itself is formulated in the full language $\mathcal{L}$; but the disquotation schema is restricted to sentences of $\mathfrak{D}$, that is, to the sentences of the base language and those needed for semantic ascent and generalization. Being a sentence of $\mathfrak{D}$ is expressed in CD by the primitive predicate symbol D, which is suitably axiomatized. As disquotation schema we can then use $\mathrm{D}\ulcorner\phi\urcorner \rightarrow (\mathrm{T}\ulcorner\phi\urcorner \leftrightarrow \phi)$ for all sentences of $\mathcal{L}$, or rather a generalization thereof with parameters. We call the sentences in $\mathfrak{D}$ *determinate*. If it were not for the additional predicate D, they would be the sentences that are grounded in Kripke's [28] sense (with some qualifications).

This approach ensures that truth can be used as a device of generalization over sentences in $\mathcal{L}_0$ and over generalizations of such generalizations, and so on. Hence this use of truth is fully available in CD. However, disquotation is not available for sentences that cannot be reached by semantic ascent and generalization.

2.1. Determinateness. We describe and axiomatize $\mathfrak{D}$, the set of determinate sentences in more detail. First, we declare all atomic sentences of the base language $\mathcal{L}_0$ determinate:

D1 $\forall s\, \forall t\, \mathrm{D}s \doteq t.$

That is, all closed identity statements are determinate. If there were further predicate symbols in the base language, analogous axioms would be added. The axioms below will enable us to prove that all sentences of the base language $\mathcal{L}_0$ are determinate.

If and only if ϕ is determinate, that is, in $\mathfrak{D}$, we add a copy $\mathrm{T}\ulcorner\phi\urcorner$ of ϕ to $\mathfrak{D}$. This is stated for arbitrary closed terms t, not only numerals $\ulcorner\phi\urcorner$:

D2 $\forall t\, \big(\mathrm{D}\dot{\mathrm{T}}t \leftrightarrow \mathrm{D}t^{\circ}\big).$

The expression t° stands for the value of the term t. We cannot have a function symbol $^{\circ}$ in our language, and thus this function needs to be expressed by a suitable formula.

A negated sentence is determinate iff the sentence is:

D4 $\forall x\, \big(\mathrm{Sent}(x) \rightarrow (\mathrm{D}(\dot{\neg}x) \leftrightarrow \mathrm{D}x)\big).$

Whether a conjunction is determinate depends on the determinateness of its conjuncts. Clearly, if both conjuncts are determinate, then so is their conjunction; and if both are indeterminate, so is their conjunction. But is a conjunction determinate if only one conjunct is?

Our choice here is motivated by the function of truth as a generalizing device. Typically, a universal generalization would have the form $\forall x\, (\phi(x) \rightarrow \mathrm{T}x)$. The formula $\phi(x)$ could express that x is a provable sentence of Peano arithmetic or that x is sentence of the form $\mathrm{T}\ulcorner\mathrm{T}\cdots\ulcorner 0=0\urcorner\urcorner\cdots\urcorner$. We consider the simple example $\forall x\, (x=\ulcorner 0=0\urcorner \rightarrow \mathrm{T}x)$ whose antecedent is satisfied by a single sentence only; we 'generalize' via semantic ascent over all sentences of the form $0=0$. Of course, generalization via semantic ascent is not needed for a single sentence; but 'generalizing' over a single sentence demonstrates the principle. If $\mathfrak{D}$ contains the sentences required for generalizing via semantic ascent, then $\forall x\, (x = \ulcorner 0=0\urcorner \rightarrow \mathrm{T}x)$ should be determinate. The determinateness of universal quantified sentences

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

depends on their instances. Let λ be an indeterminate sentence. Then the instance

$$\ulcorner\lambda\urcorner = \ulcorner 0=0\urcorner \rightarrow \mathrm{T}\ulcorner\lambda\urcorner \tag{1}$$

has a determinate antecedent and an indeterminate consequent, by axioms D1 and D2. If the instance (1) were indeterminate, so would be the universal generalization $\forall x\,(x = \ulcorner 0=0\urcorner \rightarrow \mathrm{T}x)$ by any reasonably determinateness rules for quantification. Hence, our axioms for D should declare (1) determinate.

The determinateness of (1) should follow from the determinateness of the antecedent $\ulcorner\lambda\urcorner = \ulcorner 0=0\urcorner$. The falsity of the determinate antecedent renders the consequent irrelevant; and the truth value of (1) does not depend on the consequent. Thus, we stipulate that a sentence $\phi \rightarrow \psi$ can inherit its determinateness from a determinate and false antecedent. Equally, we also postulate that it is determinate if it has a true and determinate consequent.

This method of generalization applies equally if we do not only 'generalize' over a single sentence such as $0 = 0$, but, for instance, over all sentences provable in PA. Moreover, iterated semantic ascent permits generalization over all sentences $0=0$, $\mathrm{T}\ulcorner 0=0\urcorner$, $\mathrm{T}\ulcorner \mathrm{T}\ulcorner 0=0\urcorner\urcorner$, and so on.

In our official language negation and conjunction are our only connectives. $\phi \rightarrow \psi$ is conceived as an abbreviation of $\neg(\phi \wedge \neg\psi)$. In the presence of D4 we can then state our determinateness axiom for binary connectives in the following way:

$\phi \wedge \psi$ is determinate iff:

ϕ and ψ are both determinate, or

one of the conjuncts is false and determinate.

This is expressed in the following axiom:

$$\mathsf{D5}\ \ \forall x\,\forall y\,\Big(\mathrm{Sent}(x \dot{\wedge} y) \rightarrow \big(\mathrm{D}(x \dot{\wedge} y) \leftrightarrow \big((\mathrm{D}x \wedge \mathrm{D}y) \vee (\mathrm{D}x \wedge \mathrm{F}x) \vee (\mathrm{D}y \wedge \mathrm{F}y)\big)\big)\Big).$$

By our conventions above, $\mathrm{F}x$ abbreviates $\mathrm{T}\dot{\neg}x$, namely, falsehood of x; also note that $\mathrm{D}x$ and $\mathrm{D}\dot{\neg}x$ are equivalent by the axiom D4. By axiom T4, F can be replaced with $\neg\mathrm{T}$.

Another way to argue for our treatment of binary connectives would be to argue that sentences such as $\forall x\,(x = \ulcorner 0=0\urcorner \rightarrow \mathrm{T}x)$ and its instance (1) are only about the determinate sentence $0=0$ and should be treated just like $\mathrm{T}\ulcorner 0=0\urcorner$. The latter is determinate by axiom D1. Thus, it may be argued, semantic ascent applies to (1) and the universally quantified sentence in the same way it applies to $\mathrm{T}\ulcorner 0=0\urcorner$. However, making precise the underlying notion of aboutness is notoriously challenging (see Picollo [33, 34]); we do not attempt to pursue this line here.

Universally quantified sentences are treated as conjunctions of their instances:

$$\mathsf{D6}\ \ \forall v\,\forall x\,\Big(\mathrm{Sent}(\dot{\forall} v\,x) \rightarrow \big((\mathrm{D}(\dot{\forall} v\,x) \leftrightarrow \big(\forall t\,\mathrm{D}x(t/v) \vee \exists t\big(\mathrm{D}x(t/v) \wedge \mathrm{F}x(t/v)\big)\big)\big)\Big).$$

Injecting the notion of determinateness into the object language makes it possible to apply the truth and determinateness predicates to sentences containing D. We treat both, truth and determinateness, as completely type-free: D and T can be applied meaningfully to any sentence of the language. This is in contrast to traditional conceptions of groundedness, which apply to sentences with truth, while truth cannot be applied to sentences containing groundedness claims. Therefore, in our setting, the question arises whether atomic sentences of the form $\mathrm{D}t$ are determinate.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

Treating D like an elementary predicate of $\mathcal{L}_0$, that is, a predicate from $\mathcal{L}_0$ would suggest the axiom $\forall t\, \mathrm{D}\dot{\mathrm{D}}t$. This would mean that we have $\mathrm{T}\ulcorner \mathrm{D}t \urcorner \leftrightarrow \mathrm{D}t$ for all terms t. However, when it comes to complex sentences, we axiomatize D in terms of T in D5 and D6. Hence, it looks unsafe to declare all sentence $\mathrm{D}t$ determinate; D should be treated with the same caution as T: If, and only if ϕ is determinate, we stipulate that $\mathrm{D}\ulcorner \phi \urcorner$ is determinate. Generalizing this to terms other than numerals gives the following axiom analogous to D2:

D3 $\forall t\, (\mathrm{D}\dot{\mathrm{D}}t \leftrightarrow \mathrm{D}t^{\circ})$.

This concludes the list of determinateness axioms.

2.2. Disquotation. We still need to add axioms stipulating that a sentence $\mathrm{T}\ulcorner \phi \urcorner$ in $\mathfrak{D}$ is always obtained by semantic ascent and thus a 'copy' of ϕ. Formally, we require that $\mathrm{D}\ulcorner \phi \urcorner \rightarrow (\mathrm{T}\ulcorner \phi \urcorner \leftrightarrow \phi)$ is a theorem of our theory CD for every sentence of $\mathcal{L}$. We also require that our theory proves a more general, 'uniform' version of the disquotation schema; that is, we generalize it by quantifying over the closed terms and postulate for every formula $\phi(x_1, \dots, x_n)$ with at most $x_1, \dots, x_n$ the following:

DDS $\forall t_1 \dots \forall t_n \left(\mathrm{D}\ulcorner \phi(\dot{t}_1, \dots, \dot{t}_n) \urcorner \rightarrow \left(\mathrm{T}\ulcorner \phi(\dot{t}_1, \dots, \dot{t}_n) \urcorner \leftrightarrow \phi(t_1{}^{\circ}, \dots, t_n{}^{\circ})\right)\right)$.

In this uniform determinate disquotation schema the expression $\ulcorner \phi(\dot{t}_1, \dots, \dot{t}_n) \urcorner$ stands for a complex term expressing that $t_1, \dots, t_n$ are formally substituted for the free variables in $\phi(x_1, \dots, x_n)$. This permits us to bind the variables $t_1, \dots, t_n$ in $\ulcorner \phi(\dot{t}_1, \dots, \dot{t}_n) \urcorner$. This concludes the list of axioms for our theory CD.

2.3. Alternative axiomatizations. In the presence of the compositional axioms T4 –T6, we need to stipulate DDS only for atomic ϕ. Whatever our policy on disquotation is, we can always focus on the atomic instances; T4 –T6 will ensure that also all instances of $T\ulcorner \phi \urcorner \leftrightarrow \phi$ will be provable for all sentences ϕ built from those atomic formulae. Because we aim at an axiomatization that is as lean as possible for technical reasons, our official axiomatization does not feature schema DDS, but only disquotational axioms concerning atomic sentences. In particular, schema DDS can be replaced with the following three axioms:

T1 $\forall s\, \forall t\, \left(\mathrm{T}s \doteq t \leftrightarrow s^{\circ} = t^{\circ}\right)$,
T2 $\forall t\, (\mathrm{D}t^{\circ} \rightarrow \mathrm{T}\dot{\mathrm{D}}t)$,
T3 $\forall t\, \left(\mathrm{D}t^{\circ} \rightarrow (\mathrm{T}\dot{\mathrm{T}}t \leftrightarrow \mathrm{T}t^{\circ})\right)$.

T2 can be derived from the following instance of DDS and axiom D3:

$$\forall t\, \left(\mathrm{D}\dot{\mathrm{D}}t \rightarrow (\mathrm{T}\dot{\mathrm{D}}t \leftrightarrow \mathrm{D}t^{\circ})\right). \tag{2}$$

Thus the three axioms imply DDS in the presence of the other axioms as is established in Lemma 3.1.

The principles T3, (2), and T2 show that the truth and determinateness predicates interact in a serious manner. In particular, D is not only a formalization of a metatheoretic notion that has been injected into the object language; predicates of the object language—and especially the truth predicate—can be meaningfully applied to sentences with D.

From (2) we can prove one direction of the T-sentences $\phi \rightarrow \mathrm{T}\ulcorner \phi \urcorner$ for all sentences without T, including those with D. Some sentences without T but with D—such as

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

$\mathrm{D}^{\ulcorner}\lambda^{\urcorner}$ for some indeterminate λ—are not determinate, and thus DDS will not give us the T-schema for some sentences with D. Therefore, we consider the addition of the right-to-left direction to T2:

$\mathsf{T2}^{+}\ \forall t\,(\mathrm{D}t^{\circ} \leftrightarrow \mathrm{T}\dot{\mathrm{D}}t)$.

This sentence is not provable from the disquotation schema DDS, and is not covered by our policy about $\mathfrak{D}$. With $\mathsf{T2}^{+}$ we can derive all T-sentences $\mathrm{T}^{\ulcorner}\phi^{\urcorner} \leftrightarrow \phi$ as long as ϕ is T-free.

Determinateness is one of many predicates such as analyticity, knowledge, necessity, and logical validity that are deeply intertwined with truth. Having the T-sentences $\mathrm{T}^{\ulcorner}\phi^{\urcorner} \leftrightarrow \phi$ also for ϕ containing such predicates may be desirable; but we know already that this is not be possible for all such ϕ (see Halbach [20]), as axioms like $\mathsf{T2}^{+}$ may engender inconsistency, depending on the axioms for the other predicates. We cannot think of a better strategy than to consider these axioms on a case-by-case basis and to endorse axioms like $\mathsf{T2}^{+}$, as long as they do not yield undesirable consequences. With $\mathsf{T2}^{+}$ we leave the safety of disquotation for determinate sentences only. In the case of determinateness we show that axiom $\mathsf{T2}^{+}$ does not affect the ω-soundness of our theory and thus $\mathsf{T2}^{+}$ can be added as a further optional axiom.

The quantifier $\forall t$ in D2, D3, T2, T3, and $\mathsf{T2}^{+}$ ranges over all closed terms, including those that do not denote a sentence. It may be desirable to restrict the quantifier in these axioms to terms denoting sentences. This would allow us to make any stipulations about sentences $\mathrm{T}t$ and $\mathrm{D}t$ where t fails to denote a sentence. We have chosen the version above for their simplicity.

2.4. Extensionality. In this section we discuss two additional axioms that are irrelevant to most of the metamathematical properties of our system. However, at least the first of these axioms is conceptually important. The issue goes beyond our specific system, and even beyond axiomatic theories, because semantic theories of truth are affected as well.

We expect truth to be an *extensional* notion unlike necessity, apriority, or being known. The compositional axioms prove the extensionality of truth for sentences in the following sense: Substituting a subformula in a sentence ϕ with another subformula with the same truth value does not affect truth or falsity of ϕ.

One would expect extensionality at the level of terms as well. That is, if the terms s and t have the same value, a sentence $\phi(t)$ ought to be true if, and only if $\phi(s)$ is true. However, this principle fails to be provable from the axioms we have listed so far. Only restricted versions can be proved, for instance, if the values of s and t are determinate or χ is purely arithmetical. Without axiom R1, we can refute $\mathrm{T}^{\ulcorner}\mathrm{T}s^{\urcorner} \wedge \neg\mathrm{T}^{\ulcorner}\mathrm{T}t^{\urcorner}$ under the assumption $s = t$ only for determinate s and t, but not for all terms. Here, we call a term determinate iff its value is. Thus, we state extensionality with respect to terms as an axiom:

$\mathsf{R1}\ \forall x\,\forall v\,\forall s\,\forall t\,\big(\big(\mathrm{Sent}(\dot{\forall} v\,x) \wedge s^{\circ} = t^{\circ}\big) \rightarrow \big(\mathrm{T}x(s/v) \leftrightarrow \mathrm{T}x(t/v)\big)\big)$.

Axiom R1 is closely related to the question whether identity is a logical constant. Using R1, we can prove the truth of $\forall x\,\forall y\,(x = y \rightarrow (\phi(x) \rightarrow \phi(y)))$ for all formulae ϕ, which is a logical truth if identity is a logical constant. If identity is

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

a logical constant and our theory is to prove all logically valid sentences, then we have to add R1 (or some other axiom).[3]

So far we have focused on truth. For determinateness we also stipulate extensionality:

R2 $\forall x\, \forall v\, \forall s\, \forall t\, \big((\mathrm{Sent}(\dot{\forall} v\, x) \wedge s^\circ = t^\circ) \rightarrow (\mathrm{D}x(s/v) \leftrightarrow \mathrm{D}x(t/v))\big)$.

Whether determinateness should also be extensional may be less clear. For our purposes the extensionality of determinateness is not decisive. Omitting R2 would not affect the results below.

§3. Axioms for classical determinate truth. We now collect the stipulations in the previous sections. We start from the language $\mathcal{L}_0$ of arithmetic. It contains at least a constant for 0 and a function symbol S for the successor function, a symbol $+$ for addition, and a symbol $\cdot$ for multiplication. For each number n we use the numeral $\overline{n}$, defined as the result of applying S to 0 for n-many times, as the canonical name for n. $\mathcal{L}_0$ may feature further function and constant symbols, even those for all primitive recursive functions. The identity symbol $=$ is the only predicate symbol of $\mathcal{L}_0$; but further predicate symbols could easily be added.[4]

The language $\mathcal{L}_\mathrm{T}$ is $\mathcal{L}_0$ augmented with T; $\mathcal{L}_\mathrm{D}$ is $\mathcal{L}_0$ augmented with D. Adding both T and D to $\mathcal{L}_0$ yields the language $\mathcal{L}$. Therefore, the following inclusions hold, if the languages are identified with the set of their formulae:

$$\mathcal{L}_0 \subset \mathcal{L}_\mathrm{T} \text{ and } \mathcal{L}_\mathrm{D} \subset \mathcal{L}.$$

Our system is labelled CD for 'Classical Determinate Truth'. It is formulated in the language $\mathcal{L}$ with D and T. The syntax of $\mathcal{L}$ is appropriately arithmetized within PA. We adopt the notation of Halbach [21] concerning arithmetization of syntax. However, for dealing with new predicate D and other technical reasons, we slightly supplement (and change) his notation as follows:

- $\mathrm{Sent}(x)$, $\mathrm{Sent}_\mathrm{T}(x)$, $\mathrm{Sent}_\mathrm{D}(x)$, and $\mathrm{Sent}_0(x)$ represent the set of codes of $\mathcal{L}$-, $\mathcal{L}_\mathrm{T}$-, $\mathcal{L}_\mathrm{D}$-, and $\mathcal{L}_0$-sentences, respectively.
- $\mathrm{Fml}(x)$, $\mathrm{Fml}_\mathrm{T}(x)$, $\mathrm{Fml}_\mathrm{D}(x)$, and $\mathrm{Fml}_0(x)$, respectively, represent the set of codes of formula of these languages.
- Similarly, we use $\mathrm{AtFml}(x)$, $\mathrm{AtFml}_\mathrm{T}(x)$, $\mathrm{AtFml}_\mathrm{D}(x)$, and $\mathrm{AtFml}_0(x)$ for the codes of atomic formulae of these languages, and $\mathrm{AtSent}(x)$, $\mathrm{AtSent}_\mathrm{T}(x)$, $\mathrm{AtSent}_\mathrm{D}(x)$, and $\mathrm{AtSent}_0(x)$ for the codes of atomic sentences of these languages, respectively.
- $\mathrm{Var}(x)$ and $\mathrm{Term}(x)$ represent the sets of codes of variables and codes of $\mathcal{L}$-terms (= the set of codes of $\mathcal{L}_0$-terms), respectively; recall that ClTerm is for the set of codes of *closed* $\mathcal{L}_0$-terms, and note that every closed $\mathcal{L}$-term is a closed $\mathcal{L}_0$-term.
- For each primitive predicate symbol R with arity k, $\dot{R}$ is a k-ary function that represents the syntactic operation of applying R to k-many terms. For instance,

[3] We thank Anton Broberg for pointing out to us the underivability of the truth of $\forall x\, \forall y\, (x = y \rightarrow (\phi(x) \rightarrow \phi(y)))$ in CD without axiom R1. His comment motivated the present section.

[4] With modifications the theory can reformulated in a purely relational language without closed terms such as the language of set theory. For such a setting without function symbols, a satisfaction predicate might be a better fit than a unary truth predicate (see Halbach [22]).

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

$\dot{\mathrm{D}}$ is an arithmetical representation of the syntactic operation of applying D to a term.

- $x(y/z)$ represents the syntactic substitution of a term y for a variable z in an expression x.

Hence, it can be proved in PA that $\mathrm{Sent}(\dot{\forall} v\, x)$ implies that v is the code of a variable (i.e., $\mathrm{Var}(v)$) and that x codes some formula ϕ with at most one free variable. We use $\mathrm{Sent}(\dot{\forall} v\, x)$ to express that x is a formula with at most the variable v free. Often we need to quantify over closed terms and abbreviate $\forall z\,(\mathrm{ClTerm}(z) \to \psi(z))$ as $\forall t\, \psi(t)$ and also use s as a further variable ranging over closed terms.

The system CD is given by all axioms of PA with induction in the language $\mathcal{L}$ with T and D and the following axioms:

Truth axioms

T1 $\forall s\, \forall t\, \big(\mathrm{T}s \dot{=} t \leftrightarrow s^\circ = t^\circ\big)$,
T2 $\forall t (\mathrm{D}t^\circ \to \mathrm{T}\dot{\mathrm{D}}t)$,
T3 $\forall t\, \big(\mathrm{D}t^\circ \to (\mathrm{T}\dot{\mathrm{T}}t \leftrightarrow \mathrm{T}t^\circ)\big)$,
T4 $\forall x\, \big(\mathrm{Sent}(x) \to (\mathrm{T}(\dot{\neg} x) \leftrightarrow \neg \mathrm{T}x)\big)$,
T5 $\forall x\, \forall y\, \big(\mathrm{Sent}(x \dot{\wedge} y) \to (\mathrm{T}(x \dot{\wedge} y) \leftrightarrow \mathrm{T}x \wedge \mathrm{T}y)\big)$,
T6 $\forall v\, \forall x\, \big(\mathrm{Sent}(\dot{\forall} v\, x) \to (\mathrm{T}(\dot{\forall} v\, x) \leftrightarrow \forall t\, \mathrm{T}x(t/v))\big)$.

Determinateness axioms

D1 $\forall s\, \forall t\, \mathrm{D}s \dot{=} t$,
D2 $\forall t\, \big(\mathrm{D}\dot{\mathrm{T}}t \leftrightarrow \mathrm{D}t^\circ\big)$,
D3 $\forall t\, (\mathrm{D}\dot{\mathrm{D}}t \leftrightarrow \mathrm{D}t^\circ)$,
D4 $\forall x\, \big(\mathrm{Sent}(x) \to (\mathrm{D}(\dot{\neg} x) \leftrightarrow \mathrm{D}x)\big)$,
D5 $\forall x\, \forall y\, \Big(\mathrm{Sent}(x \dot{\wedge} y) \to \big(\mathrm{D}(x \dot{\wedge} y) \leftrightarrow \big((\mathrm{D}x \wedge \mathrm{D}y) \vee (\mathrm{D}x \wedge \mathrm{F}x) \vee (\mathrm{D}y \wedge \mathrm{F}y)\big)\big)\Big)$,
D6 $\forall v\, \forall x\, \Big(\mathrm{Sent}(\dot{\forall} v\, x) \to \big((\mathrm{D}(\dot{\forall} v\, x) \leftrightarrow \big(\forall t\, \mathrm{D}x(t/v) \vee \exists t \big(\mathrm{D}x(t/v) \wedge \mathrm{F}x(t/v)\big)\big)\big)\Big)$.

As explained above, $\mathrm{F}x$ abbreviates $\mathrm{T}\dot{\neg} x$, which expresses the falsity of x.

Extensionality axioms

R1 $\forall x\, \forall v\, \forall s\, \forall t\, \big(\big(\mathrm{Sent}(\dot{\forall} v\, x) \wedge s^\circ = t^\circ\big) \to \big(\mathrm{T}x(s/v) \leftrightarrow \mathrm{T}x(t/v)\big)\big)$,
R2 $\forall x\, \forall v\, \forall s\, \forall t\, \big(\big(\mathrm{Sent}(\dot{\forall} v\, x) \wedge s^\circ = t^\circ\big) \to \big(\mathrm{D}x(s/v) \leftrightarrow \mathrm{D}x(t/v)\big)\big)$.

We now turn to some subsystems and supersystems of CD in order to assess the significance of axioms T2 and D3. It will be shown that removing one of these two axioms or both does not affect the proof-theoretic strength of CD. That is, the three systems

$$\mathsf{CD}_0 := \mathsf{CD} - \mathsf{T2} - \mathsf{D3}$$
$$\mathsf{CD}_1 := \mathsf{CD} - \mathsf{T2}$$
$$\mathsf{CD}_2 := \mathsf{CD} - \mathsf{D3}$$

are proof-theoretically equivalent, although they are different theories.

The addition of axiom $\mathsf{T2}^+$ boosts the proof-theoretic strength of CD, that is,

$$\mathsf{CD}^+ := \mathsf{CD} + \mathsf{T2}^+$$
$$\mathsf{CD}_2^+ := \mathsf{CD}_2 + \mathsf{T2}^+$$

is properly stronger than CD. Further variants of CD will be studied in Part II of the paper.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

We define disjunction $\vee$, conditional $\rightarrow$, and the existential quantifier $\exists$ in the standard manner in terms of negation $\neg$, conjunction $\wedge$, and universal quantifier $\forall$: accordingly, their corresponding operations $x \dot{\vee} y$, $\dot{\rightarrow}$, and $\dot{\exists} v x$ are defined as

$$x \dot{\vee} y := \dot{\neg}(\dot{\neg} x \dot{\wedge} \dot{\neg} y), \qquad x \dot{\rightarrow} y := \dot{\neg} x \dot{\vee} y, \quad \text{and} \quad \dot{\exists} v x := \dot{\neg}(\dot{\forall} v \dot{\neg} x).$$

It readily follows from $\mathsf{T4}$ to $\mathsf{T6}$ that T commutes with $\vee$, $\rightarrow$, and $\exists$. In addition, the following compositional rules of determinateness for these connectives are provable in CD_0:

$$\forall x \forall y \Big(\mathrm{Sent}(x \dot{\vee} y) \rightarrow \big(\mathrm{D}(x \dot{\vee} y) \leftrightarrow \big((\mathrm{D}x \wedge \mathrm{D}y) \vee (\mathrm{D}x \wedge \mathrm{T}x) \vee (\mathrm{D}y \wedge \mathrm{T}y) \big) \big) \Big), \quad (3)$$

$$\forall x \forall y \Big(\mathrm{Sent}(x \dot{\rightarrow} y) \rightarrow \big(\mathrm{D}(x \dot{\rightarrow} y) \leftrightarrow \big((\mathrm{D}x \wedge \mathrm{D}y) \vee (\mathrm{D}x \wedge \mathrm{F}x) \vee (\mathrm{D}y \wedge \mathrm{T}y) \big) \big) \Big), \quad (4)$$

$$\forall v \forall x \Big(\mathrm{Sent}(\dot{\exists} v x) \rightarrow \big(\mathrm{D}(\dot{\exists} v x) \leftrightarrow \big(\forall s \, \mathrm{D}x(s/v) \vee \exists s \big(\mathrm{D}x(s/v) \wedge \mathrm{T}x(s/v) \big) \big) \big) \Big). \quad (5)$$

Concluding this section, we mention some simple observations. First, we show that the uniform determinate disquotation schema is provable in CD, as mentioned above.

LEMMA 3.1. *For all $\mathcal{L}$-formulae $\phi(x_1, \dots, x_k)$ with at most $x_1, \dots, x_k$ free the theory* CD *proves the following*:

$$\mathsf{DDS} \quad \forall t_1 \dots \forall t_n \Big(\mathrm{D}^{\ulcorner} \phi(\underline{t}_1, \dots, \underline{t}_n)^{\urcorner} \rightarrow \big(\mathrm{T}^{\ulcorner} \phi(\underline{t}_1, \dots, \underline{t}_n)^{\urcorner} \leftrightarrow \phi(t_1^{\circ}, \dots, t_n^{\circ}) \big) \Big).$$

In particular, $\mathrm{D}^{\ulcorner} \phi^{\urcorner} \rightarrow (\mathrm{T}^{\ulcorner} \phi^{\urcorner} \leftrightarrow \phi)$ *is provable for all sentences* ϕ.

The lemma can be proved by a metatheoretic induction on the complexity of ϕ, where the complexity of a formula ϕ, $cp(\phi)$ for short, is standardly defined: every atomic formula has the complexity 0; $cp(\neg\phi) = cp(\forall x \phi) = cp(\phi) + 1$; $cp(\phi \wedge \psi) = \max\{cp(\phi), cp(\psi)\} + 1$.

By an induction within CD all sentences of the base language can be shown to be determinate. $\mathrm{Sent}_0(x)$ expresses that x is a sentence of $\mathcal{L}_0$.

LEMMA 3.2. $\mathsf{CD}_0 \vdash \forall x \, (\mathrm{Sent}_0(x) \rightarrow \mathrm{D}x)$.

Using these two lemmata, we can establish the T-sentences for all sentences of the base language.

LEMMA 3.3. *For all $\mathcal{L}_0$-formulae $\phi(x_1, \dots, x_k)$ with only $x_1, \dots, x_k$ free, we have*

$$\mathsf{CD} \vdash \forall t_1 \dots \forall t_k \, (\mathrm{T}^{\ulcorner} \phi(\underline{t}_1, \dots, \underline{t}_k)^{\urcorner} \leftrightarrow \phi(t_1^{\circ}, \dots, t_k^{\circ})).$$

Obviously, we cannot have the unrestricted T-schema in any axiomatic theory of truth. Some truth theories feature one direction of the T-schema without any restrictions. However, both directions are incompatible with CD.

PROPOSITION 3.4. *The axiom schemata* (T-*Out*) *and* (T-*In*) *are defined as follows*:

(T-Out) $\mathrm{T}^{\ulcorner} \phi^{\urcorner} \rightarrow \phi$ *for all* $\phi \in \mathcal{L}$.
(T-In) $\phi \rightarrow \mathrm{T}^{\ulcorner} \phi^{\urcorner}$ *for all* $\phi \in \mathcal{L}$.

Then, each of (T-Out) *and* (T-In) *is inconsistent with* $\mathsf{T4}$ *over arithmetic, and thus with* CD.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

PROOF. We first derive a contradiction from CD + (T-Out). Let λ be the liar sentence with $\lambda \leftrightarrow \neg \mathrm{T}\ulcorner\lambda\urcorner$ as above. The liar sentence λ implies $\mathrm{T}\ulcorner\neg\lambda\urcorner$ by T4, from which $\neg\lambda$ follows by (T-Out). Hence we have $\neg\lambda$. However, $\neg\lambda$ is equivalent to $\mathrm{T}\ulcorner\lambda\urcorner$ and thus implies λ by (T-Out).

The inconsistency of (T-In) is shown in a similar way by using the other direction of T4. ⊣

A consequence of the compositional axioms is the thorough classicality of CD. CD proves that (classical) logic is truth-preserving for the entire language $\mathcal{L}$.

LEMMA 3.5. *The following holds for all formulae $\phi(x)$ and canonical provability predicates constructed from $\phi(x)$ as defining the axioms*:

$$\mathsf{CD} \vdash \forall x\,(\phi(x) \to \mathrm{Sent}(x) \wedge \mathrm{T}x) \to \forall y\,(\mathrm{Bew}_{\phi(x)}(y) \to \mathrm{T}y).$$

Here, $\mathrm{Bew}_{\phi(x)}(y)$ expresses that y is provable from all sentences ψ with $\phi(\ulcorner\psi\urcorner)$.

PROOF. The claim is shown by induction on the length of a proof. Since T commutes with all logical connectives, quantifiers, and identity in CD, CD proves that every logical axiom is true; note that the extensionality axioms R1 and R2 are needed to show the logical axioms for identity. Every non-logical axiom is true by the assumption. The base step is thereby obtained. The induction step is straightforward, since commutativity ensures that every logical inference rule preserves truth. ⊣

Fischer et al. [12] criticized truth theories such as FS and variants of KF formulated in classical logic, because they are incompatible with reflection principles for logic. In CD the soundness of a calculus for predicate logic is provable. This is a trivial consequence of the lemma above if $\mathrm{Bew}(x)$ expresses provability in pure logic.

COROLLARY 3.6. CD *proves that every logically valid sentence (in classical logic) is true*: *namely*, $\mathsf{CD} \vdash \forall x(\mathrm{Bew}(x) \wedge \mathrm{Sent}(x) \to \mathrm{T}x)$.

§4. Semantics. McGee [31] showed that certain theories of truth that are thoroughly classical are ω-inconsistent. In particular, the system FS that also features axioms T1 and T4 –T6 is ω-inconsistent. In order to defend CD and its variants we therefore do not only show that they are consistent, but also that they possess ω-models.

In this section, we will give an ω-model of CD^+, namely, a model of CD^+ whose $\mathcal{L}_0$-reduct is the standard model of arithmetic. The existence of an ω-model entails that CD^+ and thus CD are ω-consistent.[5]

4.1. The model. Let us define $\mathcal{L}$-formulae $\mathcal{D}_i(x)$ $(1 \leq i \leq 6)$ as follows:

$\mathcal{D}_1(x) :\Leftrightarrow \exists s \exists t\,(x = (s \dot{=} t))$,
$\mathcal{D}_2(x) :\Leftrightarrow \exists s\,(x = \dot{\mathrm{T}}s \wedge \mathrm{D}s^\circ)$,
$\mathcal{D}_3(x) :\Leftrightarrow \exists s\,(x = \dot{\mathrm{D}}s \wedge \mathrm{D}s^\circ)$,
$\mathcal{D}_4(x) :\Leftrightarrow \mathrm{Sent}(x) \wedge \exists y\,(x = \dot{\neg} y \wedge \mathrm{D}y)$,

[5]The existence of an ω-model of CD^+ in itself also follows from Theorem 7.11.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

$$\mathcal{D}_5(x) :\Leftrightarrow \mathrm{Sent}(x) \wedge \exists y \exists z \left(x = y \dot{\wedge} z \wedge \left((\mathrm{D}y \wedge \mathrm{D}z) \vee (\mathrm{D}\dot{\neg} y \wedge \mathrm{F}y) \vee (\mathrm{D}\dot{\neg} z \wedge \mathrm{F}z)\right)\right),$$

$$\mathcal{D}_6(x) :\Leftrightarrow \mathrm{Sent}(x) \wedge \exists v \exists y \left(x = \dot{\forall} v\, y \wedge \left(\forall s\, \mathrm{D}y(s/v) \vee \exists s\left(\mathrm{D}\dot{\neg}(y(s/v)) \wedge \mathrm{F}(y(s/v))\right)\right)\right).$$

We thereby define $\mathcal{D}(x) :\Leftrightarrow \bigvee_{1\leq i\leq 6} \mathcal{D}_i(x)$. $\mathcal{D}$ describes the closure condition of a determinateness predicate D (relative to a fixed interpretation of the truth predicate T).

By $(\mathbb{N}, X, Y)$ let us denote the $\mathcal{L}$-structure with domain ω, the set of natural numbers, in which D is interpreted by X, T is interpreted by Y, and all the other symbols receive the standard interpretations. Both D and T occur only positively in $\mathcal{D}$. Hence, for each $Y \subset \omega$, $\mathcal{D}$ induces the following monotone operator $\Gamma_{\mathcal{D}[Y]}\colon \mathcal{P}(\omega) \to \mathcal{P}(\omega)$ so that

$$\Gamma_{\mathcal{D}[Y]}(X) = \{n \in \omega \mid (\mathbb{N}, X, Y) \models \mathcal{D}(n)\}.$$

In general, given a monotone operator $\Gamma\colon \mathcal{P}(\omega) \to \mathcal{P}(\omega)$, we say that $X \subset \omega$ is Γ-*closed* if $\Gamma(X) \subset X$, and that $X \subset \omega$ is a Γ-*fixed point* if $\Gamma(X) = X$.

Given an $\mathcal{L}_0$-formula or $\mathcal{L}_0$-term e, let us denote its standard interpretation by $e^{\mathbb{N}}$ and its Gödel number by $\#e$. For instance, $\mathrm{Sent}^{\mathbb{N}}$ denote the set of the Gödel numbers of all $\mathcal{L}$-sentences.

LEMMA 4.1. *For every $X, Y \subset \mathrm{Sent}^{\mathbb{N}}$, the following are equivalent.*[6]

1. *X is a $\Gamma_{\mathcal{D}[Y]}$-fixed point.*
2. *$(\mathbb{N}, X, Y)$ is a model of all the determinateness axioms* D1–D6.

For each ordinal α, we define sets D_α and T_α as follows:

$$D_0 := \emptyset, \qquad T_0 := \emptyset,$$

$$D_{\xi+1} := \Gamma_{\mathcal{D}[T_\xi]}(D_\xi), \qquad T_{\xi+1} := \{\#\phi \in \mathrm{Sent}^{\mathbb{N}} \mid (\mathbb{N}, D_\xi, T_\xi) \models \phi\},$$

$$D_\lambda := \bigcup_{\xi<\lambda} D_\xi, \qquad T_\lambda := \bigcup_{\xi<\lambda} T_\xi \cap D_\xi.$$

It is obvious from the definition that $T_\lambda = D_\lambda \cap T_\lambda$ for all limit ordinals λ. We can also show by induction on ordinals that $D_\xi, T_\xi \subset \mathrm{Sent}^{\mathbb{N}}$ for all ordinals ξ.

The next two propositions are obvious by definition.

LEMMA 4.2. *Let ξ be any ordinal, a be a closed $\mathcal{L}_0$-term, ϕ and ψ be $\mathcal{L}$-sentences, and $\theta(x)$ be an $\mathcal{L}$-formula with only a single variable x free.*

1. *If $\xi > 0$ and ϕ is $\mathcal{L}_0$-atomic, then $\#\phi \in D_\xi$.*
2. *$a^{\mathbb{N}} \in D_\xi$ iff $\#\mathrm{T}a \in D_{\xi+1}$.*
3. *$a^{\mathbb{N}} \in D_\xi$ iff $\#\mathrm{D}a \in D_{\xi+1}$.*

[6] $\mathcal{D}_5$ and $\mathcal{D}_6$ are defined so that the condition 1 implies that $(\mathbb{N}, X, Y)$ satisfies not the original axioms D5 and D6, but the following variants of them:

(D5') $$\forall x \forall y \left(\mathrm{Sent}(x \dot{\wedge} y) \to \left(\mathrm{D}(x \dot{\wedge} y) \leftrightarrow ((\mathrm{D}x \wedge \mathrm{D}y) \vee (\mathrm{D}\dot{\neg} x \wedge \mathrm{F}x) \vee (\mathrm{D}\dot{\neg} y \wedge \mathrm{F}y))\right)\right),$$

(D6') $$\forall v \forall x \left(\mathrm{Sent}(\dot{\forall} v\, x) \to \left(\mathrm{D}(\dot{\forall} v\, x) \leftrightarrow \left(\forall t\, \mathrm{D}x(t/v) \vee \exists t\left(\mathrm{D}\dot{\neg} x(t/v) \wedge \mathrm{F}x(t/v)\right)\right)\right)\right);$$

but they are equivalent to the original D5 and D6 anyway due to the axiom D4; we adopt this definition of $\mathcal{D}_5$ and $\mathcal{D}_6$ for a purely technical reason.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

4. $\#\phi \in D_\xi$ *iff* $\#\neg\phi \in D_{\xi+1}$.
5. $\#(\phi \wedge \psi) \in D_{\xi+1}$, *iff either* $\#\phi, \#\psi \in D_\xi$ *or* $\#\neg\phi \in D_\xi \cap T_\xi$ *or* $\#\neg\psi \in D_\xi \cap T_\xi$.
6. $\#\forall x\theta(x) \in D_{\xi+1}$, *iff either of the following holds*:
   - $\#\theta(b) \in D_\xi$ *for all* $\mathcal{L}_0$*-closed terms* b;
   - $\#\neg\theta(b) \in D_\xi \cap T_\xi$ *for some* $\mathcal{L}_0$*-closed term* b.

PROOF. For example, we have $\#\mathrm{T}a = \dot{\mathrm{T}}^{\mathbb{N}}(\#a)$ and $((\#a)^\circ)^{\mathbb{N}} = a^{\mathbb{N}}$, from which claim 2 follows, since we have

$$a^{\mathbb{N}} \in D_\xi \quad \Leftrightarrow \quad (\mathbb{N}, D_\xi, T_\xi) \models \mathcal{D}_2(\#\mathrm{T}a) \quad \Leftrightarrow \quad \#\mathrm{T}a \in D_{\xi+1}. \qquad \dashv$$

LEMMA 4.3. *Let* ξ *be any ordinal,* a *and* b *be closed* $\mathcal{L}_0$*-terms,* ϕ *and* ψ *be* $\mathcal{L}$*-sentences, and* $\theta(x)$ *be an* $\mathcal{L}$*-formula only with some single variable* x *free.*

1. $\#(a = b) \in T_{\xi+1}$ *iff* $a^{\mathbb{N}} = b^{\mathbb{N}}$.
2. $a^{\mathbb{N}} \in T_\xi$ *iff* $\#\mathrm{T}a \in T_{\xi+1}$.
3. $a^{\mathbb{N}} \in D_\xi$ *iff* $\#\mathrm{D}a \in T_{\xi+1}$.
4. $\#\neg\phi \in T_{\xi+1}$ *iff* $\phi \notin T_{\xi+1}$.
5. $\#(\phi \wedge \psi) \in T_{\xi+1}$ *iff* $\#\phi \in T_{\xi+1}$ *and* $\#\psi \in T_{\xi+1}$.
6. $\#\forall x\theta(x) \in T_{\xi+1}$ *iff* $\#\theta(c) \in T_{\xi+1}$ *for all closed* $\mathcal{L}_0$*-terms* c.

Let us define a formula $x \approx y$ as follows:

$$x \approx y \ :\Leftrightarrow\ x = y \ \vee\ \exists z\, \exists v\, \exists s\, \exists t\, \big(\mathrm{Sent}(\forall v\, z) \wedge s^\circ{=}t^\circ \wedge x = z(s/v) \wedge y = z(t/v)\big).$$

Namely, for natural numbers $n, m \in \mathbb{N}$, $n \approx^{\mathbb{N}} m$ means that either $n = m$ or n and m code numerically equivalent $\mathcal{L}$-sentences in the sense that they are substitution instances of the same $\mathcal{L}$-formula with closed terms with the same values. Since all the existential quantifiers in the definition can actually be bounded by the values of some primitive recursive functions on x and y, the relation $\approx^{\mathbb{N}}$ is primitive recursive *in* the value function (or the index function $[\,\cdot\,]$ of the primitive recursive functions) and thus provably recursive in PA. The extensionality axioms R1 and R2 say that truth and determinateness are invariant across numerically equivalent $\mathcal{L}$-sentences, namely, sentences ϕ and ψ with $\#\phi \approx^{\mathbb{N}} \#\psi$.[7] The following proposition is shown by a straightforward induction on ordinals.

PROPOSITION 4.4. *Let* ξ *be any ordinal and* $n \approx^{\mathbb{N}} m$.

1. $n \in D_\xi$ *iff* $m \in D_\xi$.
2. $n \in T_\xi$ *iff* $m \in T_\xi$; *note that if* $\#\phi \approx^{\mathbb{N}} \#\psi$ *then* $(\mathbb{N}, X, Y) \models \phi \leftrightarrow \psi$.

The next is the main lemma of this subsection.

LEMMA 4.5. *For all ordinals* ξ *and* η *with* $\xi \leq \eta$, *the following hold*:

1. $D_\xi \cap T_\xi = D_\xi \cap T_\eta$ *and*
2. $D_\xi \subset D_\eta$.

[7] The first disjunct '$x = y$' is added for a purely technical reason, namely, for incorporating the ordinary logical axioms for atomic sentences of the form $\mathrm{D}a$ or $\mathrm{T}a$ to the axioms (Ax2) and (Ax3) of the semi-formal system CD^∞, which will be introduced later in Section 7.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

PROOF. By simultaneous induction on ξ. The base step is trivial. Let ξ be a limit ordinal and $\eta \geq \xi$. For each $n \in D_\xi$, there is some $\zeta < \xi$ such that $n \in D_\zeta$, for which we have $D_\zeta \cap T_\xi = D_\zeta \cap T_\zeta = D_\zeta \cap T_\eta$ and $D_\zeta \subset D_\eta$ by the induction hypothesis.

Let ξ be a successor ordinal. Claim 1 is shown by sub-induction on η; we can assume $\eta > \xi$. First, suppose η is a limit. If $n \in D_\xi \cap T_\xi$, then $n \in T_\eta$ by definition. Let $n \in D_\xi \cap T_\eta$ for the converse. By definition, there is some $\zeta < \eta$ such that $n \in D_\zeta \cap T_\zeta$. If $\zeta < \xi$, then $D_\zeta \cap T_\zeta \subset T_\xi$ by the induction hypothesis for 1; otherwise, $\xi \leq \zeta < \eta$ and thus $D_\xi \cap T_\zeta = D_\xi \cap T_\xi$ by the sub-induction hypothesis. Next, suppose η is a successor, and take any $n = \#\phi \in D_\xi$ $(\subset \mathrm{Sent}^{\mathbb{N}})$ for an $\mathcal{L}$-sentence ϕ. If ϕ is an $\mathcal{L}_0$-atomic, then the claim follows from Lemmata 4.2.1 and 4.3.1. If $\phi = \mathrm{T}a$ for some closed $\mathcal{L}_0$-term a, then we have $a^{\mathbb{N}} \in D_{\xi-1}$ by Lemma 4.2.2, and thus we obtain

$$n \in T_\xi \overset{4.3.2}{\Leftrightarrow} a^{\mathbb{N}} \in T_{\xi-1} \overset{\mathrm{IH}}{\Leftrightarrow} a^{\mathbb{N}} \in T_{\eta-1} \overset{4.3.2}{\Leftrightarrow} n \in T_\eta,$$

where 'IH' denotes the induction hypothesis. If $\phi = \mathrm{D}a$, then we have $a^{\mathbb{N}} \in D_{\xi-1}$ by Lemma 4.2.3, and thus $n \in T_\xi$ and $a^{\mathbb{N}} \in D_{\eta-1}$ by Lemma 4.3.3 and the induction hypothesis for 2, respectively, the latter of which implies $n \in T_\eta$ again by Lemma 4.3.3; hence, both $n \in T_\xi$ and $n \in T_\eta$ hold. We move on to the cases for complex ϕs. Let $\phi = \psi \wedge \theta$. By Lemma 4.2.5, $n \in D_\xi$ implies

$$\big(\#\psi, \#\theta \in D_{\xi-1}\big) \vee \big(\#\neg\psi \in D_{\xi-1} \cap T_{\xi-1}\big) \vee \big(\#\neg\theta \in D_{\xi-1} \cap T_{\xi-1}\big).$$

If the first disjunct holds, then we have

$$n \in T_\xi \overset{4.3.5}{\Leftrightarrow} \#\psi, \#\theta \in T_\xi \overset{\mathrm{IH}}{\Leftrightarrow} \#\psi, \#\theta \in T_{\xi-1} \overset{\mathrm{IH}}{\Leftrightarrow} \#\psi, \#\theta \in T_\eta \overset{4.3.5}{\Leftrightarrow} n \in T_\eta.$$

If the second disjunct holds, then we have $\#\neg\psi \in T_\xi \cap T_\eta$ by the induction hypothesis for 1, from which we can infer

$$\#\neg\psi \in T_\xi \cap T_\eta \overset{4.3.4}{\Leftrightarrow} \#\psi \notin T_\xi \text{ and } \#\psi \notin T_\eta \overset{4.3.5}{\Rightarrow} \#\phi \notin T_\xi \text{ and } \#\phi \notin T_\eta.$$

The case where the third disjunct holds can be similarly treated. The claim for the remaining cases where ϕ is of the form $\neg\psi$ or $\forall x \psi(x)$ can be shown similarly.

Claim 2 for a successor ξ is also shown by sub-induction on η. As before, we can assume $\eta > \xi$, and the limit case is obvious. Let η be a successor ordinal and take any $n = \#\phi \in D_\xi$ for an $\mathcal{L}$-sentence ϕ. If ϕ is $\mathcal{L}_0$-atomic, then the claim trivially obtains by Lemma 4.2.1. If either $\varphi = \mathrm{T}a$ or $\varphi = \mathrm{D}a$, then we have

$$n \in D_\xi \overset{4.2.2 \text{ or } 3}{\Leftrightarrow} a^{\mathbb{N}} \in D_{\xi-1} \overset{\mathrm{IH}}{\Rightarrow} a^{\mathbb{N}} \in D_{\eta-1} \overset{4.2.2 \text{ or } 3}{\Leftrightarrow} n \in D_\eta.$$

The claim for the other cases for complex ϕs can be straightforwardly shown. For instance, if $\phi = \psi \wedge \theta$, then we have

$$\begin{aligned} n \in D_\xi &\overset{4.2.5}{\Leftrightarrow} \big(\#\psi, \#\theta \in D_{\xi-1}\big) \vee \big(\#\neg\psi \in D_{\xi-1} \cap T_{\xi-1}\big) \vee \big(\#\neg\theta \in D_{\xi-1} \cap T_{\xi-1}\big) \\ &\overset{\mathrm{IH}}{\Rightarrow} \big(\#\psi, \#\theta \in D_{\eta-1}\big) \vee \big(\#\neg\psi \in D_{\eta-1} \cap T_{\eta-1}\big) \vee \big(\#\neg\theta \in D_{\eta-1} \cap T_{\eta-1}\big) \\ &\overset{4.2.5}{\Leftrightarrow} n \in D_\eta. \end{aligned}$$

The other cases can be treated in a similar way. ⊣

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

LEMMA 4.6. *Let ϕ be an $\mathcal{L}$-sentence and ξ an ordinal.*
1. *If $\#\phi \in D_\xi$, then $\#\phi \in T_\xi$ iff $(\mathbb{N}, D_\xi, T_\xi) \models \phi$.*
2. *If $a^{\mathbb{N}} = \#\phi \in D_\xi$, then $(\mathbb{N}, D_\xi, T_\xi) \models \phi$ iff $(\mathbb{N}, D_\xi, T_\xi) \models \mathrm{T}a$.*

PROOF. 1. Let $\#\phi \in D_\xi$. Then, we have

$$\#\phi \in T_\xi \overset{4,5.1}{\Leftrightarrow} \#\phi \in T_{\xi+1} \overset{\text{def}}{\Leftrightarrow} \#(\mathbb{N}, D_\xi, T_\xi) \models \phi.$$

2. Let $a^{\mathbb{N}} = \#\phi \in D_\xi$. Then, we have

$$(\mathbb{N}, D_\xi, T_\xi) \models \phi \overset{\text{by 1}}{\Leftrightarrow} \#\phi \in T_\xi \overset{\text{def}}{\Leftrightarrow} (\mathbb{N}, D_\xi, T_\xi) \models \mathrm{T}a. \qquad \dashv$$

LEMMA 4.7. 1. *For each ξ and $\mathcal{L}$-sentence ϕ, either $\#\phi \notin T_\xi$ or $\#\neg\phi \notin T_\xi$.*
2. *For each ξ and $\mathcal{L}$-sentence ϕ, if $\#\phi \in D_\xi$, then either $\#\phi \in T_\xi$ or $\#\neg\phi \in T_\xi$.*

PROOF. 1. By induction on ξ. The claim for a non-limit ξ immediately follows by definition. Let ξ be a limit, and suppose for contradiction that $\#\phi \in T_\xi$ and $\#\neg\phi \in T_\xi$. Then, there are some $\zeta, \eta < \xi$ such that $\#\phi \in T_\zeta \cap D_\zeta$ and $\#\neg\phi \in T_\eta \cap D_\eta$, which is impossible by Lemma 4.5 and the induction hypothesis.

2. By induction on ξ. If $\xi = 0$, then $D_0 = \emptyset$ and the claim trivially holds. If ξ is a successor, then the claim immediately follows by Lemma 4.3.4. Finally, if ξ is a limit and $\#\phi \in D_\xi$, then $\#\phi \in D_\zeta$ for some $\zeta < \xi$, and thus either $\#\phi \in T_\zeta$ or $\neg\phi \in T_\zeta$ by the induction hypothesis, from which the claim follows by Lemma 4.5.1. $\dashv$

Let ω_1 denote the least uncountable ordinal. By Lemma 4.5, we have $D_\xi \subset D_\eta$ and $D_\xi \cap T_\xi \subset D_\eta \cap T_\eta$ for all $\xi \leq \eta$. Hence, by the standard cardinality consideration, we have the following nice properties of D_{ω_1} and T_{ω_1}:

$$D_{\omega_1+1} = D_{\omega_1}, \text{ and } D_{\omega_1+1} \cap T_{\omega_1+1} = D_{\omega_1} \cap T_{\omega_1} = T_{\omega_1}; \tag{6}$$

in particular, D_{ω_1} is a $\Gamma_{\mathcal{D}[T_{\omega_1}]}$-fixed point. Let us denote D_{ω_1} and T_{ω_1} by D_∞ and T_∞, respectively. We finally give an ω-model of CD.

THEOREM 4.8. *Let $\mathbb{T}_\infty := \{\#\phi \mid (\mathbb{N}, D_\infty, T_\infty) \models \phi\}$. Then, $(\mathbb{N}, D_\infty, \mathbb{T}_\infty) \models \mathsf{CD}^+$. Hence, in particular, CD^+ and CD have an ω-model and thus are ω-consistent.*

PROOF. By definition, it immediately follows that $(\mathbb{N}, D_\infty, \mathbb{T}_\infty)$ is a model of $\mathsf{T1}$, $\mathsf{T2}^+$, and $\mathsf{T4}$–$\mathsf{T6}$. To see $(\mathbb{N}, D_\infty, \mathbb{T}_\infty) \models \mathsf{T3}$, take any $n \in \mathrm{ClTerm}^{\mathbb{N}}$ and suppose $(n^\circ)^{\mathbb{N}} \in D_\infty$. Then, there exist some closed $\mathcal{L}_0$-term a and $\mathcal{L}$-sentence ϕ such that $n = \#a$ and $(n^\circ)^{\mathbb{N}} = a^{\mathbb{N}} = \#\phi \in D_\infty$ ($\subset \mathrm{Sent}^{\mathbb{N}}$). Hence, we have

$$\begin{aligned}(\mathbb{N}, D_\infty, \mathbb{T}_\infty) \models \mathrm{T}\dot{\mathrm{T}}n \overset{\text{def}}{\Leftrightarrow} (\mathbb{N}, D_\infty, T_\infty) \models \mathrm{T}a &\overset{4.6.2}{\Leftrightarrow} (\mathbb{N}, D_\infty, T_\infty) \models \phi \\ &\overset{\text{def}}{\Leftrightarrow} (\mathbb{N}, D_\infty, \mathbb{T}_\infty) \models \mathrm{T}n^\circ.\end{aligned}$$

Since D_∞ is a $\Gamma_{\mathcal{D}[T_\infty]}$-fixed point, $(\mathbb{N}, D_\infty, \mathbb{T}_\infty)$ satisfies $\mathsf{D1}$–$\mathsf{D4}$ (which does not depend on the interpretation of T). To show $(\mathbb{N}, D_\infty, \mathbb{T}_\infty) \models \mathsf{D5}$, take any $\mathcal{L}$-sentences ϕ and ψ. By Lemma 4.6.1, we have $D_\infty \cap \mathbb{T}_\infty = D_\infty \cap T_\infty$ and thus obtain

$$\begin{aligned}\#\phi \wedge \psi \in D_\infty &\overset{4.2.5}{\Leftrightarrow} (\#\phi, \#\psi \in D_\infty) \vee (\#\neg\phi \in D_\infty \cap T_\infty) \vee (\#\neg\psi \in D_\infty \cap T_\infty) \\ &\Leftrightarrow (\#\phi, \#\psi \in D_\infty) \vee (\#\neg\phi \in D_\infty \cap \mathbb{T}_\infty) \vee (\#\neg\psi \in D_\infty \cap \mathbb{T}_\infty).\end{aligned}$$

We can similarly show $(\mathbb{N}, D_\infty, \mathbb{T}_\infty) \models \mathsf{D6}$ using $D_\infty \cap T_\infty = D_\infty \cap \mathbb{T}_\infty$. $\dashv$

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

4.2. The minimality of D. In this section, we will show that D_∞ ($:= D_{\omega_1}$) has a certain 'minimality' property.

Let us define $\mathcal{L}$-formulae $\mathcal{T}_j(x)$ $(1 \leq j \leq 6)$ as follows:

$$\mathcal{T}_1(x) :\Leftrightarrow \exists s \exists t \Big(\big(x = (s \doteq t) \wedge s^\circ = t^\circ\big) \vee \big(x = (\dot\neg s \doteq t) \wedge s^\circ \neq t^\circ\big)\Big),$$

$$\mathcal{T}_2(x) :\Leftrightarrow \exists s \Big(\big(x = \dot{\mathrm{T}} s \wedge \mathrm{D} s^\circ \wedge \mathrm{T} s^\circ\big) \vee \big(x = \dot\neg \dot{\mathrm{T}} s \wedge \mathrm{D}(\dot\neg s^\circ) \wedge \mathrm{F} s^\circ\big)\Big),$$

$$\mathcal{T}_3(x) :\Leftrightarrow \exists s \, (x = \dot{\mathrm{D}} s \wedge \mathrm{D} s^\circ),$$

$$\mathcal{T}_4(x) :\Leftrightarrow \mathrm{Sent}(x) \wedge \exists y \, (x = \dot\neg \dot\neg y \wedge \mathrm{T} y),$$

$$\mathcal{T}_5(x) :\Leftrightarrow \mathrm{Sent}(x) \wedge \exists y \exists z \Big(\big(x = (y \dot\wedge z) \wedge \mathrm{T} y \wedge \mathrm{T} z\big) \vee \big(x = \dot\neg(y \dot\wedge z) \wedge (\mathrm{F} y \vee \mathrm{F} z)\big)\Big),$$

$$\mathcal{T}_6(x) :\Leftrightarrow \mathrm{Sent}(x) \wedge \exists v \exists y \Big(\big(x = \dot\forall v \, y \wedge \forall s \mathrm{T} y(s/v)\big) \vee \big(x = \dot\neg \dot\forall v \, y \wedge \exists s \mathrm{F} y(s/v)\big)\Big).$$

We define $\mathcal{T}(x) :\Leftrightarrow \bigvee_{1 \leq j \leq 6} \mathcal{T}_j(x)$. Note that $\mathcal{T}$ says nothing about the Gödel numbers of $\mathcal{L}$-sentences of the form $\neg \mathrm{D} a$. Both D and T occur only positively in $\mathcal{T}$. Hence, for each given $X \subset \omega$, it induces a monotone operator $\Gamma_{\mathcal{T}[X]} \colon \mathcal{P}(\omega) \to \mathcal{P}(\omega)$ such that

$$\Gamma_{\mathcal{T}[X]}(Y) = \{n \in \omega \mid (\mathbb{N}, X, Y) \models \mathcal{T}(n)\};$$

note that $\Gamma_{\mathcal{T}[X]}(Y) \subset \mathrm{Sent}^{\mathbb{N}}$ for all $X, Y \subset \omega$.

LEMMA 4.9. *Let $X, Y \subset \omega$. If $(\mathbb{N}, X, Y) \models \mathsf{CD}$, then Y is $\Gamma_{\mathcal{T}[X]}$-closed.*

PROOF. The proof is routine. Take any $n \in \Gamma_{\mathcal{T}[X]}(Y) \subset \mathrm{Sent}^{\mathbb{N}}$ and let $n = \#\phi$ where ϕ is an $\mathcal{L}$-sentence. If $\phi = \mathrm{D} a$ for a closed $\mathcal{L}_0$-term a, then we have $a^{\mathbb{N}} \in X$ and thus $(\mathbb{N}, X, Y) \models \mathrm{D} a^\circ$, which implies $(\mathbb{N}, X, Y) \models \mathrm{T}\dot{\mathrm{D}} a$, namely, $\#\phi \in Y$, because $(\mathbb{N}, X, Y) \models \mathsf{T2}$; note that it is never the case that $\#\phi = \neg \mathrm{D} a$ for any closed $\mathcal{L}_0$-term a. If $\phi = \neg \mathrm{T} a$ for a closed $\mathcal{L}_0$-term a, then $(\dot\neg a)^{\mathbb{N}} \in X \cap Y$ and thus $a^{\mathbb{N}} \in X \setminus Y$ by $(\mathbb{N}, X, Y) \models \mathsf{D4} \wedge \mathsf{T4}$, which implies $\#\phi \in Y$ because $(\mathbb{N}, X, Y) \models \mathsf{T3} \wedge \mathsf{T4}$. Next, let $\phi = \neg(\psi \wedge \theta)$. Then, either $\#\neg\psi \in Y$ or $\#\neg\theta \in Y$. If the former is the case, then $\#\psi \notin Y$ by $(\mathbb{N}, X, Y) \models \mathsf{T4}$ and thus $\#\phi \notin Y$ by $(\mathbb{N}, X, Y) \models \mathsf{T5}$, from which we get $\#\neg\phi \in Y$ by $(\mathbb{N}, X, Y) \models \mathsf{T4}$. We leave the remaining cases to the reader. ⊣

LEMMA 4.10. *Let $X, Y \subset \omega$. Suppose the following*:

(a) *X is $\Gamma_{\mathcal{D}[Y]}$-closed*;
(b) *Y is $\Gamma_{\mathcal{T}[X]}$-closed.*

Then we have $D_\infty \subset X$ and $T_\infty \subset Y$ and thus, by (6) above, $T_\infty \subset D_\infty \cap Y$.

PROOF. It suffices to show the following under the supposition of (a) and (b): for all $\mathcal{L}$-sentences ϕ and ordinals α,

$$\#\phi \in D_\alpha \to \#\phi \in X, \tag{7}$$

$$\#\phi \in D_\alpha \cap T_\alpha \to \#\phi \in Y, \tag{8}$$

$$\#\phi \in D_\alpha \setminus T_\alpha \to \#\neg\phi \in Y. \tag{9}$$

They are shown by simultaneous induction on α. The base step is trivial. If α is a limit, the claim follows from the induction hypothesis using Lemma 4.5. Let α be

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

a successor. The claim for the case where ϕ is an $\mathcal{L}_0$-atomic is obvious. We will go through the remaining cases.

Let $\phi = \mathrm{T}a$ and suppose $\#\phi \in D_\alpha$. We have $a^{\mathbb{N}} \in D_{\alpha-1}$ ($\subset \mathrm{Sent}^{\mathbb{N}}$) by Lemma 4.2.2. Then, we have $a^{\mathbb{N}} \in X$ by the induction hypothesis and thus $\#\phi \in X$ by (a); hence (7) holds for ϕ. The other claims (8) and (9) are obtained as follows:

$$\#\phi \in T_\alpha \overset{4.3.2}{\Leftrightarrow} a^{\mathbb{N}} \in T_{\alpha-1} \overset{\mathrm{IH}}{\Rightarrow} a^{\mathbb{N}} \in X \cap Y \overset{(\mathrm{b})}{\Rightarrow} \#\phi \in Y;$$

$$\#\phi \notin T_\alpha \overset{4.3.2}{\Leftrightarrow} a^{\mathbb{N}} \notin T_{\alpha-1} \overset{\mathrm{IH}}{\Rightarrow} (\dot{\neg}a)^{\mathbb{N}} \in X \cap Y \overset{(\mathrm{b})}{\Rightarrow} \#\neg\phi \in Y.$$

Let $\phi = \mathrm{D}a$ and suppose $\#\phi \in D_\alpha$. We have $a^{\mathbb{N}} \in D_{\alpha-1}$ by Lemma 4.2.3. Then, we get $a^{\mathbb{N}} \in X$ by the induction hypothesis, which implies $\#\phi \in Y$ by (b); hence, (7) and (8) hold for ϕ. The other claim (9) trivially holds, since $\#\phi \notin T_\alpha$ would imply $a^{\mathbb{N}} \notin D_{\alpha-1}$ by Lemma 4.3.3, which is absurd.

Let $\phi = \neg\psi$ and suppose $\#\phi \in D_\alpha$. We have $\#\psi \in D_{\alpha-1}$ by Lemma 4.2.4. Then, it follows that $\#\psi \in X$ by the induction hypothesis and thus $\#\phi \in X$ by (a). The other claims (8) and (9) are obtained as follows:

$$\#\phi \in T_\alpha \overset{4.3.4}{\Leftrightarrow} \#\psi \notin T_\alpha \overset{4.5.1}{\Leftrightarrow} \#\psi \notin T_{\alpha-1} \overset{\mathrm{IH}}{\Rightarrow} \#\phi \in Y;$$

$$\#\phi \notin T_\alpha \overset{4.3.4}{\Leftrightarrow} \#\psi \in T_\alpha \overset{4.5.1}{\Leftrightarrow} \#\psi \in T_{\alpha-1} \overset{\mathrm{IH}}{\Rightarrow} \#\psi \in Y \overset{(\mathrm{b})}{\Rightarrow} \#\neg\phi \in Y.$$

Let $\phi = \psi \wedge \theta$ and suppose $\#\phi \in D_\alpha$. By Lemma 4.2.5, there are three cases to be separately considered. First assume $\#\psi, \#\theta \in D_{\alpha-1}$. Then, we get $\#\psi \in X$ and $\#\theta \in X$ by the induction hypothesis, and thus $\#\phi \in X$ by (a); hence, (7) holds for ϕ. We next obtain (8) as follows:

$$\#\phi \in T_\alpha \overset{4.3.5}{\Leftrightarrow} \#\psi, \#\theta \in T_\alpha \overset{4.5.1}{\Leftrightarrow} \#\psi, \#\theta \in T_{\alpha-1} \overset{\mathrm{IH}}{\Rightarrow} \#\psi, \#\theta \in Y \overset{(\mathrm{b})}{\Rightarrow} \#\phi \in Y.$$

For the remaining claim (9), we infer

$$\#\phi \notin T_\alpha \overset{4.3.5}{\Leftrightarrow} \#\psi \notin T_\alpha \text{ or } \#\theta \notin T_\alpha \overset{4.5.1}{\Leftrightarrow} \#\psi \notin T_{\alpha-1} \text{ or } \#\theta \notin T_{\alpha-1}$$
$$\overset{\mathrm{IH}}{\Rightarrow} \#\neg\psi \in Y \text{ or } \#\neg\theta \in Y \overset{(\mathrm{b})}{\Rightarrow} \#\neg\phi \in Y.$$

Second assume $\#\neg\psi \in D_{\alpha-1} \cap T_{\alpha-1}$. We have $\#\neg\psi \in X \cap Y$ by the induction hypothesis, which implies $\#\phi \in X$ and $\#\neg\phi \in Y$ by (a) and (b); hence, (7) and (9) hold for ϕ. Note that $\#\phi \in T_\alpha$ can never be the case under the current assumption, since it would imply $\#\neg\psi \notin T_\alpha$ by Lemmata 4.3.4 and 4.3.5 and thus $\#\neg\psi \notin T_{\alpha-1}$ by Lemma 4.5.1; hence, (8) trivially holds for ϕ. The case where $\#\neg\theta \in D_{\alpha-1} \cap T_{\alpha-1}$ can be similarly treated.

The remaining case where $\phi = \forall x\psi(x)$ can be similarly treated to the last case, and we omit the details. ⊣

Remark 4.11. As we have remarked, D_∞ is $\Gamma_{\mathcal{D}[T_\infty]}$-closed. We can also show that T_∞ is $\Gamma_{\mathcal{T}[D_\infty]}$-closed; the proof is routine. Hence, it follows from Lemma 4.10 that D_∞ and T_∞ are simultaneously inductively defined by the following monotone operators from $\mathcal{P}(\mathbb{N}) \times \mathcal{P}(\mathbb{N})$ to $\mathcal{P}(\mathbb{N})$:

$$\Gamma_{\mathcal{D}}(X, Y) = \Gamma_{\mathcal{D}[Y]}(X) \qquad \text{and} \qquad \Gamma_{\mathcal{T}}(X, Y) = \Gamma_{\mathcal{T}[X]}(Y).$$

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

LEMMA 4.12. *Let $X, Y \subset \omega$. Suppose the same conditions (a) and (b) as in Lemma 4.10, as well as the following additional condition*:

(c) for all $\mathcal{L}$-sentences ϕ, if $\#\phi \in X$, then either $\#\phi \notin Y$ or $\#\neg\phi \notin Y$.

Then the following holds.

1. $D_\infty \cap Y = T_\infty$.
2. D_∞ *is the least* $\Gamma_{\mathcal{D}[Y]}$*-closed set*.

PROOF. We have $T_\infty \subset D_\infty$ by (6) and $T_\infty \subset Y$ by Lemma 4.10; hence, $T_\infty \subset D_\infty \cap Y$. For the converse, take any $\#\phi \notin T_\infty$. If $\#\phi \in D_\infty$, then we have $\#\neg\phi \in T_\infty \subset Y$ by Lemma 4.7.2 and thus $\#\phi \notin Y$ by (c).

For claim 2, we first show that D_∞ is $\Gamma_{\mathcal{D}[Y]}$-closed. We observe that T only appears in the clauses $\mathcal{D}_5$ and $\mathcal{D}_6$ of $\mathcal{D}$ in the form $(\mathrm{D}\dot{\neg}x \wedge \mathrm{T}\dot{\neg}x)$. Hence, since $D_\infty \cap Y = D_\infty \cap T_\infty$ by claim 1, we have $\Gamma_{\mathcal{D}[Y]}(D_\infty) = \Gamma_{\mathcal{D}[T_\infty]}(D_\infty) \subset D_\infty$. Now, take any $\Gamma_{\mathcal{D}[Y]}$-closed set Z. Since the intersection of $\Gamma_{\mathcal{D}[Y]}$-closed sets is also $\Gamma_{\mathcal{D}[Y]}$-closed (by the monotonicity of $\Gamma_{\mathcal{D}[Y]}$), $D_\infty \cap Z$ is $\Gamma_{\mathcal{D}[Y]}$-closed. Also, since D occurs in $\mathcal{T}$ only positively, it also follows that Y is $\Gamma_{\mathcal{T}[D_\infty \cap Z]}$-closed; for, $D_\infty \subset X$ by Lemma 4.10 and thus $\Gamma_{\mathcal{T}[D_\infty \cap Z]}(Y) \subset \Gamma_{\mathcal{T}[D_\infty]}(Y) \subset \Gamma_{\mathcal{T}[X]}(Y) \subset Y$. Hence, $D_\infty \cap Z$ and Y satisfy the conditions (a) and (b) of Lemma 4.10, and thus we obtain $D_\infty \subset Z$. ⊣

The next theorem immediately follows from Lemmata 4.1, 4.9, and 4.12.

THEOREM 4.13. *Let $X, Y \subset \mathbb{N}$. If $(\mathbb{N}, X, Y) \models \mathsf{CD}$, then the following hold*:

1. $D_\infty \subset X$.
2. $D_\infty \cap Y = T_\infty$.
3. D_∞ *is the least* $\Gamma_{\mathcal{D}[Y]}$*-fixed point*.

This theorem says that every sentence in D_∞ is determinate in any ω-model of CD, the truth and the falsity are invariable on D_∞ across all ω-models of CD, and D_∞ is the least fixed point of $\Gamma_{\mathcal{D}[Y]}$ (and thus is the least set Z with $(\mathbb{N}, Z, Y) \models \mathrm{D1} - \mathrm{D6}$) whenever $(\mathbb{N}, X, Y) \models \mathsf{CD}$ for some X. This mathematical fact, as well as the philosophical story that motivated CD, suggests an axiom expressing such a minimality property of D, which yields a new theory CD_μ. This axiom and the resulting system will be briefly explained in Section 9, but their full analysis is left for the Part II.

§5. A lower bound of the strength of CD. In this section, we will show that the system $\mathsf{RA}_{<\varepsilon_0}$ of ramified analysis up to ε_0 and the system $\mathsf{RT}_{<\varepsilon_0}$ of ε_0-iterated truth (see [21] for its definition) are arithmetically conservative over CD, that is, every arithmetical theorem of the former system is provable in the latter system, and also that $\mathsf{RA}_{<\varepsilon_{\varepsilon_0}}$ and $\mathsf{RT}_{<\varepsilon_{\varepsilon_0}}$ are arithmetically conservative over CD^+. The reductions will be achieved by constructing a relative interpretation of intermediate systems KF and CT[[KF]] in CD and CD^+, respectively.

5.1. The systems KF and CT. For the sake of the reader's convenience, we repeat the definition of the system KF of Feferman [8] (in our notation).

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

DEFINITION 5.1. The $\mathcal{L}_{\mathrm{T}}$-system KF is defined as PA with full induction in $\mathcal{L}_{\mathrm{T}}$ augmented with the following axioms:

KF1 $\forall s \forall t \left(\mathrm{T}(s \doteq t) \leftrightarrow s^{\circ} = t^{\circ}\right) \wedge \forall s \forall t \left(\mathrm{F}(s \doteq t) \leftrightarrow s^{\circ} \neq t^{\circ}\right)$,
KF2 $\forall s \left(\mathrm{T}\dot{\mathrm{T}}s \leftrightarrow \mathrm{T}s^{\circ}\right) \wedge \forall s \left(\mathrm{F}\dot{\mathrm{T}}s \leftrightarrow \mathrm{F}s^{\circ}\right)$,
KF3 $\forall x \left(\mathrm{Sent}_{\mathrm{T}}(x) \rightarrow \left(\mathrm{T}(\dot{\neg}\dot{\neg}x) \leftrightarrow \mathrm{T}x\right)\right)$,
KF4 $\forall x \forall y \left(\mathrm{Sent}_{\mathrm{T}}(x \dot{\wedge} y) \rightarrow \left(\mathrm{T}(x \dot{\wedge} y) \leftrightarrow (\mathrm{T}x \wedge \mathrm{T}y)\right)\right)$,
KF5 $\forall x \forall y \left(\mathrm{Sent}_{\mathrm{T}}(x \dot{\wedge} y) \rightarrow \left(\mathrm{F}(x \dot{\wedge} y) \leftrightarrow (\mathrm{F}x \vee \mathrm{F}y)\right)\right)$,
KF6 $\forall v \forall x \left(\mathrm{Sent}_{\mathrm{T}}(\dot{\forall} v\, x) \rightarrow \left(\mathrm{T}(\dot{\forall} v\, x) \leftrightarrow \forall s\, \mathrm{T}x(s/v)\right)\right)$,
KF7 $\forall v \forall x \left(\mathrm{Sent}_{\mathrm{T}}(\dot{\forall} v\, x) \rightarrow \left(\mathrm{F}(\dot{\forall} v\, x) \leftrightarrow \exists s \mathrm{F}x(s/v)\right)\right)$,
KF8 $\forall x\, \forall v\, \forall s\, \forall t \left(\left(\mathrm{Sent}_{\mathrm{T}}(\dot{\forall} v x) \wedge s^{\circ} = t^{\circ}\right) \rightarrow \left(\mathrm{T}x(s/v) \leftrightarrow \mathrm{T}x(t/v)\right)\right)$.

Note that, while $\mathrm{F}x$ and $\neg \mathrm{T}x$ are equivalent in KD, they are not equivalent in KF, and we have to distinguish them when working in KF. It is easily shown that KF8 is redundant and provable from the other axioms (see [5, Lemma 3.1]). We will occasionally consider two extra axioms Cons and Comp, which are defined as follows:

Cons: $\forall x \left(\mathrm{Sent}_{\mathrm{T}}(x) \rightarrow \neg(\mathrm{T}x \wedge \mathrm{F}x)\right)$,

Comp: $\forall x \left(\mathrm{Sent}_{\mathrm{T}}(x) \rightarrow (\mathrm{T}x \vee \mathrm{F}x)\right)$.

For each $\varphi \in \mathcal{L}_{\mathrm{T}}$, let φ^{c} denote Cantini's 'dual' translation of φ: namely, $\mathrm{T}^{c}a := \neg \mathrm{F}a$ for each $\mathcal{L}_0$-term a, and c preserves the $\mathcal{L}_0$-vocabulary (as well as the logical connectives and quantifiers); see [5, Section 4]. c is a relative interpretation of KF + Cons in KF + Comp and *vice versa*. Cantini [5] showed that KF + Cons and KF + Comp are both proof-theoretically equivalent to KF.

LEMMA 5.2. KF *proves the following.*

1. $\forall x \forall y \left(\mathrm{Sent}_{\mathrm{T}}(x \dot{\vee} y) \rightarrow \left(\left(\mathrm{T}(x \dot{\vee} y) \leftrightarrow (\mathrm{T}x \vee \mathrm{T}y)\right) \wedge \left(\mathrm{F}(x \dot{\vee} y) \leftrightarrow (\mathrm{F}x \wedge \mathrm{F}y)\right)\right)\right)$.
2. $\forall x \forall y \left(\mathrm{Sent}_{\mathrm{T}}(x \dot{\rightarrow} y) \rightarrow \left(\left(\mathrm{T}(x \dot{\rightarrow} y) \leftrightarrow (\neg \mathrm{F}x \rightarrow \mathrm{T}y)\right) \wedge \left(\mathrm{F}(x \dot{\rightarrow} y) \leftrightarrow (\mathrm{T}x \wedge \mathrm{F}y)\right)\right)\right)$.
3. $\forall v \forall x \left(\dot{\forall} v x \in \mathrm{Sent}_{\mathrm{T}} \rightarrow \left(\left(\mathrm{T}(\dot{\exists} v x) \leftrightarrow \exists s \mathrm{T}x(s/v)\right) \wedge \mathrm{F}(\dot{\exists} v x) \leftrightarrow \forall s \mathrm{F}x(s/v)\right)\right)$.

DEFINITION 5.3. Let $\mathcal{L}_1$ be a first-order language that extends $\mathcal{L}_0$ only with new predicate symbols. We set a new language $\mathcal{L}_1^{+}$ to be $\mathcal{L}_1 \cup \{\mathbb{T}\}$ where $\mathbb{T}$ is a fresh unary predicate symbol: the new predicate $\mathbb{T}$ will be interpreted as the Tarskian typed compositional truth for the language $\mathcal{L}_1$. Let Sent_1 be a representation of the set of (codes of) $\mathcal{L}_1$-sentences. Given any $\mathcal{L}_1$-system S, the $\mathcal{L}_1^{+}$-system CT[S] is defined as S with all axiom schemata of S (possibly including other schemata than induction) extended for $\mathcal{L}_1^{+}$ together with the following axioms expressing Tarski's 'inductive clauses' of truth.

$\mathbb{T}$1 $\forall \vec{s} \left(\mathbb{T}\dot{\mathrm{R}}\vec{s} \leftrightarrow \mathrm{R}\vec{s^{\circ}}\right)$, for all $\mathcal{L}_1$-atomic formulae $\mathrm{R}\vec{x}$.
$\mathbb{T}$2 $\forall x \left(\mathrm{Sent}_1(x) \rightarrow \left(\mathbb{T}(\dot{\neg} x) \leftrightarrow \neg \mathbb{T}x\right)\right)$.
$\mathbb{T}$3 $\forall x \forall y \left(\mathrm{Sent}_1(x \dot{\wedge} y) \rightarrow \left(\mathbb{T}(x \dot{\wedge} y) \leftrightarrow \mathbb{T}x \wedge \mathbb{T}y\right)\right)$.
$\mathbb{T}$4 $\forall v \forall x \left(\dot{\forall} v x \in \mathrm{Sent}_1 \rightarrow \left(\mathbb{T}(\dot{\forall} v x) \leftrightarrow \forall s\, \mathbb{T}x(s/v)\right)\right)$.
R4 $\forall x\, \forall v\, \forall s\, \forall t \left(\left(\mathrm{Sent}_1(\dot{\forall} v x) \wedge s^{\circ} = t^{\circ}\right) \rightarrow \left(\mathbb{T}x(s/v) \leftrightarrow \mathbb{T}x(t/v)\right)\right)$.

Here we abbreviate a sequence of variables (or terms), $x_0, \ldots, x_k$ (or $a_1, \ldots, a_k$), by $\vec{x}$ (or $\vec{a}$) for saving space. If $\mathcal{L}_1 \supset \mathcal{L}_{\mathrm{T}}$, then $\mathbb{T}$1 contains $\forall s \left(\mathbb{T}\dot{\mathrm{T}}s \leftrightarrow \mathrm{T}s^{\circ}\right)$. It can be

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

shown in a parallel manner to the case of R3 in KF that R4 is derivable from the other axioms.

LEMMA 5.4. *Let* S *be as above.* CT[S] *proves the following theorems*:

1. $\forall x\, \forall y \left(\mathrm{Sent}_1(x \dot{\vee} y) \rightarrow \left(\mathbb{T}(x \dot{\vee} y) \leftrightarrow (\mathbb{T}x \vee \mathbb{T}y)\right)\right)$.
2. $\forall x\, \forall y \left(\mathrm{Sent}_1(x \dot{\rightarrow} y) \rightarrow \left(\mathbb{T}(x \dot{\rightarrow} y) \leftrightarrow (\mathbb{T}x \rightarrow \mathbb{T}y)\right)\right)$.
3. $\forall v\, \forall x \left(\mathrm{Sent}_1(\dot{\forall} vx) \rightarrow \left(\mathbb{T}(\dot{\exists} vx) \leftrightarrow \exists s \mathbb{T}x(s/v)\right)\right)$.

5.2. Reduction of KF to CD_0. The translation $\natural$ of $\mathcal{L}_\mathrm{T}$ into $\mathcal{L}$ replaces every atomic formula $\mathrm{T}^\natural a$ (for a term a) with $\mathrm{T}a \wedge \mathrm{D}a$; $\natural$ does not change anything else.

In what follows, we will occasionally write $s \in \mathrm{ClTerm}$ instead of $\mathrm{ClTerm}(s)$, $x \in \mathrm{Sent}$ instead of $\mathrm{Sent}(x)$, and similarly for other formulae.

LEMMA 5.5. $\natural$ *is a relative interpretation of* KF *in* CD_0.

PROOF. We will only exhibit the proofs that $\natural$ preserves the axioms K2, K5, and K7. For K2, take any $s \in \mathrm{ClTerm}$. Then we have

$$\mathrm{T}\dot{\neg}\dot{\mathrm{T}}s \wedge \mathrm{D}\dot{\neg}\dot{\mathrm{T}}s \overset{\mathsf{T4}\,\&\,\mathsf{D4}}{\Leftrightarrow} \neg\mathrm{T}\dot{\mathrm{T}}s \wedge \mathrm{D}\dot{\mathrm{T}}s \overset{\mathsf{D2}}{\Leftrightarrow} \neg\mathrm{T}\dot{\mathrm{T}}s \wedge \mathrm{D}s^\circ \overset{\mathsf{T3}}{\Leftrightarrow} \neg\mathrm{T}s^\circ \wedge \mathrm{D}s^\circ \overset{\mathsf{T4}\,\&\,\mathsf{D4}}{\Leftrightarrow} \mathrm{T}\dot{\neg}s^\circ \wedge \mathrm{D}\dot{\neg}s^\circ.$$

The other conjunct can be shown similarly. For K5, take any $x, y \in \mathrm{Sent}_\mathrm{T}$; then we have

$$\begin{aligned}
\mathrm{T}\dot{\neg}(x \dot{\wedge} y) \wedge \mathrm{D}\dot{\neg}(x \dot{\wedge} y) &\overset{\mathsf{T4}\,\&\,\mathsf{T5}}{\Leftrightarrow} (\mathrm{F}x \vee \mathrm{F}y) \wedge \mathrm{D}\dot{\neg}(x \dot{\wedge} y) \\
&\overset{\mathsf{D4}\,\&\,\mathsf{D5}}{\Leftrightarrow} (\mathrm{F}x \vee \mathrm{F}y) \wedge \left((\mathrm{D}x \wedge \mathrm{D}y) \vee (\mathrm{D}x \wedge \mathrm{F}x) \vee (\mathrm{D}y \wedge \mathrm{F}y)\right) \\
&\overset{\text{by logic}}{\Leftrightarrow} (\mathrm{D}x \wedge \mathrm{F}x) \vee (\mathrm{D}y \wedge \mathrm{F}y) \\
&\overset{\mathsf{D4}}{\Leftrightarrow} (\mathrm{D}\dot{\neg}x \wedge \mathrm{T}\dot{\neg}x) \vee (\mathrm{D}\dot{\neg}y \wedge \mathrm{T}\dot{\neg}y).
\end{aligned}$$

For K7, take any v and x such that $\dot{\forall} vx \in \mathrm{Sent}_\mathrm{T}$; then we have

$$\begin{aligned}
\mathrm{T}\dot{\neg}(\dot{\forall} vx) \wedge \mathrm{D}\dot{\neg}(\dot{\forall} vx) &\overset{\mathsf{T4}\,\&\,\mathsf{T6}}{\Leftrightarrow} \exists s\, \mathrm{F}x(s/v) \wedge \mathrm{D}\dot{\neg}(\dot{\forall} vx) \\
&\overset{\mathsf{D4}\,\&\,\mathsf{D6}}{\Leftrightarrow} \exists s\, \mathrm{F}x(s/v) \wedge \left(\forall s \mathrm{D}x(s/v) \vee \exists s\, (\mathrm{D}x(s/v) \wedge \mathrm{F}x(s/v))\right) \\
&\overset{\text{by logic}}{\Leftrightarrow} \exists s\, (\mathrm{D}x(s/v) \wedge \mathrm{F}x(s/v)) \\
&\overset{\mathsf{D4}}{\Leftrightarrow} \exists s\, (\mathrm{D}\dot{\neg}x(s/v) \wedge \mathrm{T}\dot{\neg}x(s/v)).
\end{aligned}$$

The remaining cases are similarly and even more easily shown. ⊣

Because $\natural$ does not affect arithmetical sentences, the next corollary follows.

COROLLARY 5.6. KF *is arithmetically conservative over* CD_0 *and thus* CD.

Finally, since KF and $\mathsf{RA}_{<\varepsilon_0}$ are known to have the same arithmetical theorems by results due to Cantini [5] and Feferman [8], $\mathsf{RA}_{<\varepsilon_0}$ is arithmetically conservative over CD.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

5.3. Reduction of $\mathsf{CT}[\![\mathsf{KF}]\!]$ to CD_2^+. Let k be an $\mathcal{L}_0$-definable function that represents the arithmetization of the translation $\natural$ of $\mathcal{L}_\mathrm{T}$ in $\mathcal{L}$: that is, for all $x \in \omega$,

$$kx := \begin{cases} x, & \text{if } x \in \mathrm{AtFml}_0, \\ (\dot{\mathrm{T}}y)\dot{\wedge}(\dot{\mathrm{D}}y), & \text{if } x = \dot{\mathrm{T}}y \text{ and } y \in \mathrm{Term}, \\ \dot{\neg}ky, & \text{if } x = \dot{\neg}y \in \mathrm{Fml}_\mathrm{T}, \\ (ky)\dot{\wedge}(kz), & \text{if } x = y\dot{\wedge}z \in \mathrm{Fml}_\mathrm{T}, \\ \dot{\forall}v(ky), & \text{if } x = \dot{\forall}vy \in \mathrm{Fml}_\mathrm{T}, \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

With the standard coding, k is primitive recursive and thus definable in PA, and PA proves $\forall x\big(\mathrm{Fml}_\mathrm{T}(x) \to \mathrm{Fml}(kx)\big)$. Furthermore, with the notation of [5], PA also proves $\forall z\big(\mathrm{Free}(z,x) \leftrightarrow \mathrm{Free}(z,kx)\big)$, where $\mathrm{Free}(a,b)$ says that a is a code of a variable that is free in the formula coded by b; in particular, we have $\forall x\big(\mathrm{Sent}_\mathrm{T}(x) \leftrightarrow \mathrm{Sent}(kx)\big)$.

We thereby extend $\natural$ to a translation of $\mathcal{L}_\mathrm{T}^+$ to $\mathcal{L}$ by stipulating the following:

$$\mathbb{T}^\natural x := \mathrm{T}kx.$$

We will use the same symbol $\natural$ for the two translations for simplicity.

LEMMA 5.7. $\mathsf{PA} \vdash (\forall x \in \mathrm{Fml}_\mathrm{T})\forall s(\forall v \in \mathrm{Var})\big(k(x(s/v)) = (kx)(s/v)\big)$.

PROOF. By a routine induction on the complexity of x. ⊣

LEMMA 5.8. *The following claims are provable in* CD_2.

1. $\forall x\,\big(\mathrm{Sent}_\mathrm{T}(x) \to \big(\mathrm{T}k(\dot{\neg}x) \leftrightarrow \neg \mathrm{T}kx\big)\big)$; *by* T4 *and* (10).
2. $\forall x\,\forall y\,\big(\mathrm{Sent}_\mathrm{T}(x\dot{\wedge}y) \to \big(\mathrm{T}k(x\dot{\wedge}y) \leftrightarrow (\mathrm{T}kx \wedge \mathrm{T}ky)\big)\big)$; *by* T5 *and* (10).
3. $\forall x\,\forall v\,\big(\mathrm{Sent}_\mathrm{T}(\dot{\forall}v\,x) \to \big(\mathrm{T}k(\dot{\forall}vx) \leftrightarrow \forall s\,\mathrm{T}k(x(s/v))\big)\big)$; *by* T6, (10), *and Lemma* 5.7.

THEOREM 5.9. $\natural$ *is a relative interpretation of* $\mathsf{CT}[\![\mathsf{KF}]\!]$ *in* CD_2^+.

PROOF. It immediately follows from the definition of $\natural$ and the last lemma that the $\natural$-translations of the axioms $\mathbb{T}1$ for all $\mathcal{L}_0$-atomics and $\mathbb{T}2$–$\mathbb{T}4$ are provable in CD_2. $\mathsf{R4}^\natural$ readily follows from R1 and Lemma 5.7 (though this step is actually redundant because R4 is derivable from the other axioms of $\mathsf{CT}[\![\mathsf{KF}]\!]$). For the remaining case, i.e., the axiom $\mathbb{T}1$ for an atomic of the form $\mathrm{T}x$, take any $s \in \mathrm{ClTerm}$. Then we have

$$\mathbb{T}^\natural(\dot{\mathrm{T}}s) \Leftrightarrow \mathrm{T}(\dot{\mathrm{T}}s\dot{\wedge}\dot{\mathrm{D}}s) \overset{\mathsf{T5}}{\Leftrightarrow} \mathrm{T}\dot{\mathrm{T}}s \wedge \mathrm{T}\dot{\mathrm{D}}s \overset{\mathsf{T2}^+}{\Leftrightarrow} \mathrm{T}\dot{\mathrm{T}}s \wedge \mathrm{D}s^\circ \overset{\mathsf{T3}}{\Leftrightarrow} \mathrm{T}s^\circ \wedge \mathrm{D}s^\circ \Leftrightarrow \mathrm{T}^\natural s^\circ;$$

note that the use of $\mathsf{T2}^+$ in the third equivalence is crucial here. ⊣

COROLLARY 5.10. $\mathsf{CT}[\![\mathsf{KF}]\!]$ *is arithmetically conservative over* CD_2^+ *and thus* CD^+.

Now, the following fact is essentially due to Cantini [5].

FACT 5.11. $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!]$ $(= \mathsf{CT}[\![\mathsf{KF}]\!] + \mathrm{Cons})$ *is arithmetically equivalent to* $\mathsf{RA}_{<\varepsilon_{\varepsilon_0}}$, *that is, the two systems have the same arithmetical theorems.*

Hence, $\mathsf{RT}_{<\varepsilon_{\varepsilon_0}}$ is arithmetically conservative over CD^+.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

§6. An upper bound of the strength of CD. In this section, we will show that CD is proof-theoretically reducible to KF. This will be achieved via partial cut-elimination for a semi-formal system for CD.

6.1. Semi-formal system CD^∞. In this subsection, we will introduce a Tait-style semi-formal system CD^∞.

Throughout this Section 6, capital Greek letters Γ and Δ will be used as variables ranging over finite sets of $\mathcal{L}$-sentences, and we will only consider $\mathcal{L}$-sentences in negation-normal form (or in what is sometimes called 'Tait form'), which are constructed from closed $\mathcal{L}_0$-literals by conjunction, disjunction, and universal and existential quantifiers; we will abuse the notation and denote the negation operation on negation-normal forms, which is often written as '$\sim$' in the literature, also by $\neg$, e.g., $\neg(\mathrm{T}a \wedge \mathrm{D}a) = \neg\mathrm{T}a \vee \neg\mathrm{D}a$. Following the convention, for a finite set Γ of $\mathcal{L}$-sentences and an $\mathcal{L}$-sentence A, we will write Γ, A for $\Gamma \cup \{A\}$.

The derivation relation $\mathsf{CD}^\infty \vdash^{\delta}_{p} \Gamma$ in the semi-formal system CD^∞ for ordinals δ and $p < \omega$ and a set Γ of $\mathcal{L}$-sentences, which intuitively means that at least one $\phi \in \Gamma$ is verified by a derivation of length δ with cut-rank p, is defined as follows.

Axioms. Let a, b, and c be any closed $\mathcal{L}_0$-terms, Γ any set of $\mathcal{L}$-sentences, δ any ordinal, and $p < \omega$ any finite ordinal.

(Ax1) If A is a true closed $\mathcal{L}_0$-literal, then $\mathsf{CD}^\infty \vdash^{\delta}_{p} \Gamma, A$.

(Ax2) If $a^{\mathbb{N}} \approx^{\mathbb{N}} b^{\mathbb{N}}$ (see p.13 for the definition of $\approx$), then $\mathsf{CD}^\infty \vdash^{\delta}_{p} \Gamma, \mathrm{D}a, \neg\mathrm{D}b$.

(Ax3) If $a^{\mathbb{N}} \approx^{\mathbb{N}} b^{\mathbb{N}}$, then $\mathsf{CD}^\infty \vdash^{\delta}_{p} \Gamma, \mathrm{T}a, \neg\mathrm{T}b$.

(D1) If $b^{\mathbb{N}}, c^{\mathbb{N}} \in \mathrm{ClTerm}^{\mathbb{N}}$ and $a^{\mathbb{N}} = (b \doteq c)^{\mathbb{N}}$, then $\mathsf{CD}^\infty \vdash^{\delta}_{p} \Gamma, \mathrm{D}a$.

Logical Rules. Let A and B be $\mathcal{L}$-sentences, $C(x)$ an $\mathcal{L}$-formula with at most one free variable, δ an ordinal, and $p < \omega$.

($\vee$1) If $\mathsf{CD}^\infty \vdash^{\delta'}_{p} \Gamma, A \vee B, A$ for some ordinal $\delta' < \delta$, then $\mathsf{CD}^\infty \vdash^{\delta}_{p} \Gamma, A \vee B$.

($\vee$2) If $\mathsf{CD}^\infty \vdash^{\delta'}_{p} \Gamma, A \vee B, B$ for some ordinal $\delta' < \delta$, then $\mathsf{CD}^\infty \vdash^{\delta}_{p} \Gamma, A \vee B$.

($\bigwedge$) If $\mathsf{CD}^\infty \vdash^{\delta_0}_{p} \Gamma, A \wedge B, A$ and $\mathsf{CD}^\infty \vdash^{\delta_1}_{p} \Gamma, A \wedge B, B$ for some ordinals $\delta_0, \delta_1 < \delta$, then $\mathsf{CD}^\infty \vdash^{\delta}_{p} \Gamma, A \wedge B$.

($\exists$) If $\mathsf{CD}^\infty \vdash^{\delta'}_{p} \Gamma, \exists x C(x), C(a)$ for some closed $\mathcal{L}_0$-term a and ordinal $\delta' < \delta$, then $\mathsf{CD}^\infty \vdash^{\delta}_{p} \Gamma, \exists x C(x)$.

($\forall$) If there exists some ordinal $\delta_a < \delta$ for each closed $\mathcal{L}_0$-term a such that $\mathsf{CD}^\infty \vdash^{\delta_a}_{p} \Gamma, \forall x C(x), C(a)$, then $\mathsf{CD}^\infty \vdash^{\delta}_{p} \Gamma, \forall x C(x)$.

(cut) If $\mathsf{CD}^\infty \vdash^{\delta_0}_{p} \Gamma, A$ and $\mathsf{CD}^\infty \vdash^{\delta_1}_{p} \Gamma, \neg A$ for some ordinals $\delta_0, \delta_1 < \delta$ and $\mathcal{L}$-sentence A with complexity $< p$, then $\mathsf{CD}^\infty \vdash^{\delta}_{p} \Gamma$.

Rules for positive occurrences of D. For ordinals δ and $p < \omega$ and a closed term a, $\mathsf{CD}^\infty \vdash^{\delta}_{p} \Gamma, \mathrm{D}a$ holds, if one of the following holds for some ordinal $\delta' < \delta$[8]:

[8]Note that $\mathsf{D5}^{\pm}_{\infty}$ and $\mathsf{D6}^{\pm}_{\infty}$ are the rule versions of the variants D5' and D6' of D5 and D6 (see fn. 4.1 for their definitions). Recall that they are equivalent to the original D5 and D6 due to the axiom D4.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

($\mathrm{D2}^{pos}_{\infty}$) There is a closed term b such that $a^{\mathbb{N}} = (\dot{\mathrm{T}}b)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^{\infty} \vdash^{\delta'}_{p} \Gamma, \mathrm{D}a, \mathrm{D}b^{\circ}$.

($\mathrm{D3}^{pos}_{\infty}$) There is a closed term b such that $a^{\mathbb{N}} = (\dot{\mathrm{D}}b)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^{\infty} \vdash^{\delta'}_{p} \Gamma, \mathrm{D}a, \mathrm{D}b^{\circ}$.

($\mathrm{D4}^{pos}_{\infty}$) There is a closed term b such that $a^{\mathbb{N}} = (\dot{\neg}b)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^{\infty} \vdash^{\delta'}_{p} \Gamma, \mathrm{D}a, \mathrm{D}b$.

($\mathrm{D5}^{pos}_{\infty}$) There are some closed terms b and c such that $a^{\mathbb{N}} = (b \dot{\wedge} c)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^{\infty} \vdash^{\delta'}_{p} \Gamma, \mathrm{D}a, (\mathrm{D}b \wedge \mathrm{D}c) \vee (\mathrm{D}\dot{\neg}b \wedge \mathrm{F}b) \vee (\mathrm{D}\dot{\neg}c \wedge \mathrm{F}c)$.

($\mathrm{D6}^{pos}_{\infty}$) There are some closed terms b and c such that $a^{\mathbb{N}} = (\dot{\forall} bc)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^{\infty} \vdash^{\delta'}_{p} \Gamma, \mathrm{D}a, \forall s \mathrm{D}c(s/b) \vee \exists s(\mathrm{D}\dot{\neg}c(s/b) \wedge \mathrm{F}c(s/b))$.

Rules for negative occurrences of D. For ordinals δ and $p < \omega$ and a closed term a, $\mathsf{CD}^{\infty} \vdash^{\delta}_{p} \Gamma, \neg\mathrm{D}a$ holds, if one of the following holds for some ordinal $\delta' < \delta$:

($\mathrm{D2}^{neg}_{\infty}$) There is a closed term b such that $a^{\mathbb{N}} = (\dot{\mathrm{T}}b)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^{\infty} \vdash^{\delta'}_{p} \Gamma, \neg\mathrm{D}a, \neg\mathrm{D}b^{\circ}$.

($\mathrm{D3}^{neg}_{\infty}$) There is a closed term b such that $a^{\mathbb{N}} = (\dot{\mathrm{D}}b)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^{\infty} \vdash^{\delta'}_{p} \Gamma, \neg\mathrm{D}a, \neg\mathrm{D}b^{\circ}$.

($\mathrm{D4}^{neg}_{\infty}$) There is a closed term b such that $a^{\mathbb{N}} = (\dot{\neg}b)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^{\infty} \vdash^{\delta'}_{p} \Gamma, \neg\mathrm{D}a, \neg\mathrm{D}b$.

($\mathrm{D5}^{neg}_{\infty}$) There are some closed terms b and c such that $a^{\mathbb{N}} = (b \dot{\wedge} c)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^{\infty} \vdash^{\delta'}_{p} \Gamma, \neg\mathrm{D}a, (\neg\mathrm{D}b \vee \neg\mathrm{D}c) \wedge (\neg\mathrm{D}\dot{\neg}b \vee \neg\mathrm{F}b) \wedge (\neg\mathrm{D}\dot{\neg}c \vee \neg\mathrm{F}c)$.

($\mathrm{D6}^{neg}_{\infty}$) There are some closed terms b and c such that $a^{\mathbb{N}} = (\dot{\forall} bc)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^{\infty} \vdash^{\delta'}_{p} \Gamma, \neg\mathrm{D}a, \exists s \neg\mathrm{D}c(s/b) \wedge \forall s(\neg\mathrm{D}\dot{\neg}c(s/b) \vee \neg\mathrm{F}c(s/b))$.

Rules for positive occurrences of T. For ordinals δ and $p < \omega$ and a closed term a, $\mathsf{CD}^{\infty} \vdash^{\delta}_{p} \Gamma, \mathrm{T}a$ holds, if one of the following holds for some ordinal $\delta' < \delta$:

($\mathrm{T1}^{pos}_{\infty}$) There are some closed terms b and c such that $a^{\mathbb{N}} = (b \dot{=} c)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^{\infty} \vdash^{\delta'}_{p} \Gamma, \mathrm{T}a, b^{\circ} = c^{\circ}$.

($\mathrm{T2}^{pos}_{\infty}$) There is a closed term b such that $a^{\mathbb{N}} = (\dot{\mathrm{D}}b)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^{\infty} \vdash^{\delta'}_{p} \Gamma, \mathrm{T}a, \mathrm{D}b^{\circ}$.

($\mathrm{T3}^{pos}_{\infty}$) There is a closed term b such that $a^{\mathbb{N}} = (\dot{\mathrm{T}}b)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^{\infty} \vdash^{\delta'}_{p} \Gamma, \mathrm{T}a, \mathrm{T}b^{\circ} \wedge \mathrm{D}b^{\circ}$.

($\mathrm{T4}^{pos}_{\infty}$) There is a closed term b such that $a^{\mathbb{N}} = (\dot{\neg}b)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^{\infty} \vdash^{\delta'}_{p} \Gamma, \mathrm{T}a, \neg\mathrm{T}b$.

($\mathrm{T5}^{pos}_{\infty}$) There are some closed terms b and c such that $a^{\mathbb{N}} = (b \dot{\wedge} c)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^{\infty} \vdash^{\delta'}_{p} \Gamma, \mathrm{T}a, \mathrm{T}b \wedge \mathrm{T}c$.

($\mathrm{T6}^{pos}_{\infty}$) There are some closed terms b and c such that $a^{\mathbb{N}} = (\dot{\forall} bc)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^{\infty} \vdash^{\delta'}_{p} \Gamma, \mathrm{T}a, \forall s \mathrm{T}c(s/b)$.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

Rules for a negative occurrence of T. For ordinals δ and $p < \omega$ and a closed term a, $\mathsf{CD}^\infty \vdash^\delta_p \Gamma, \neg \mathrm{T}a$ holds, if one of the following holds for some ordinal $\delta' < \delta$:

- $(\mathrm{T}1^{neg}_\infty)$ There are some closed terms b and c such that $a^{\mathbb{N}} = (b \doteq c)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^\infty \vdash^{\delta'}_p \Gamma, \neg \mathrm{T}a, b^\circ \neq c^\circ$.
- $(\mathrm{T}3^{neg}_\infty)$ There is a closed term b such that $a^{\mathbb{N}} = (\dot{\mathrm{T}}b)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^\infty \vdash^{\delta'}_p \Gamma, \neg \mathrm{T}a, \neg \mathrm{T}b^\circ \wedge \mathrm{D}b^\circ$.
- $(\mathrm{T}4^{neg}_\infty)$ There is a closed term b such that $a^{\mathbb{N}} = (\dot{\neg} b)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^\infty \vdash^{\delta'}_p \Gamma, \neg \mathrm{T}a, \mathrm{T}b$.
- $(\mathrm{T}5^{neg}_\infty)$ There are some closed terms b and c such that $a^{\mathbb{N}} = (b \dot{\wedge} c)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^\infty \vdash^{\delta'}_p \Gamma, \neg \mathrm{T}a, \neg \mathrm{T}b \vee \neg \mathrm{T}c$.
- $(\mathrm{T}6^{neg}_\infty)$ There are some closed terms b and c such that $a^{\mathbb{N}} = (\dot{\forall} bc)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$ and $\mathsf{CD}^\infty \vdash^{\delta'}_p \Gamma, \mathrm{T}a, \exists s \neg \mathrm{T}c(s/b)$.

Note that there is no rule for a negative occurrence of T corresponding to $(\mathrm{T}2^{pos}_\infty)$.

LEMMA 6.1. *The following basic properties of* CD^∞ *can be shown in the standard manner, and we omit their proofs.*

1. (*Structural Lemma*)
   For all $\gamma \leq \delta$, $q \leq p$, *and* $\Delta \subset \Gamma$, *if* $\mathsf{CD}^\infty \vdash^\gamma_q \Delta$, *then* $\mathsf{CD}^\infty \vdash^\delta_p \Gamma$.
2. (*Numerical Equivalence Lemma*)
   For all closed terms a *and* b, *if* $a^{\mathbb{N}} = b^{\mathbb{N}}$ *and* $\mathsf{CD}^\infty \vdash^\delta_p \Gamma, A(a)$, *then* $\mathsf{CD}^\infty \vdash^\delta_p \Gamma, A(b)$.
3. (*Tautology Lemma*)
   $\mathsf{CD}^\infty \vdash^{2 \cdot cp(A)}_0 A, \neg A$; *recall that* $cp(A)$ *denotes the complexity of* A (*see page* 12).
4. ($\wedge$-*Inversion*)
   If $\mathsf{CD}^\infty \vdash^\delta_p \Gamma, A \wedge B$, *then* $\mathsf{CD}^\infty \vdash^\delta_p \Gamma, A$, *and* $\mathsf{CD}^\infty \vdash^\delta_p \Gamma, B$.
5. ($\vee$-*Exportation*)
   If $\mathsf{CD}^\infty \vdash^\delta_p \Gamma, A \vee B$, *then* $\mathsf{CD}^\infty \vdash^\delta_p \Gamma, A, B$.
6. ($\forall$-*Inversion*)
   If $\mathsf{CD}^\infty \vdash^\delta_p \Gamma, \forall x A(x)$, *then* $\mathsf{CD}^\infty \vdash^\delta_p \Gamma, A(a)$ *for all closed terms* a.

The next lemma is nearly obvious from the design of CD^∞, and the proof is routine.

LEMMA 6.2. *For every* $\mathcal{L}$*-sentence* ψ, *if* $\mathsf{CD} \vdash \psi$, *then* $\mathsf{CD}^\infty \vdash^{\omega+k}_n \psi$ *for some* $k, n < \omega$; *actually, we can primitive recursively calculate* k *and* n *from a given derivation of* ψ.

The following three lemmata can also be shown standardly.

LEMMA 6.3 (Reduction Lemma). *If* $\mathsf{CD}^\infty \vdash^\alpha_{1+n+1} \Gamma$, *then* $\mathsf{CD}^\infty \vdash^{2^\alpha}_{1+n} \Gamma$.

LEMMA 6.4 (Partial Cut-elimination Lemma). *If* $\mathsf{CD}^\infty \vdash^\alpha_{1+n} \Gamma$, *then* $\mathsf{CD}^\infty \vdash^{\omega_n(\alpha)}_1 \Gamma$, *where* $\omega_n(\alpha)$ *is recursively defined as follows*: $\omega_0(\alpha) = \alpha$ *and* $\omega_{k+1}(\alpha) = \omega^{\omega_k(\alpha)}$.

LEMMA 6.5. *For every* $\mathcal{L}$*-sentence* ψ, *if* $\mathsf{CD} \vdash \psi$, *then* $\mathsf{CD}^\infty \vdash^\delta_1 \psi$ *for some* $\delta < \varepsilon_0$.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

Note that $\mathsf{CD}^\infty \vdash^\delta_1 \psi$ means that there is a derivation of ψ in CD^∞ in which (cut) is only applied to closed $\mathcal{L}$-literals; such a derivation corresponds to what Cantini [5] calls a *quasi-normal* derivation in his semi-formal system TKF^∞.

6.2. Reduction of CD to KF. Let σ, τ, and ξ be ordinals. Recall that we only consider $\mathcal{L}$-sentences in negation-normal form throughout the Section 6. Given an $\mathcal{L}$-sentence ϕ, we write $\models \phi[\sigma, \tau, \xi]$ when ϕ is true under the following interpretation:

- the quantifiers range over ω;
- the $\mathcal{L}_0$-vocabulary receives the standard interpretations;
- each negative occurrence of D in ϕ is interpreted by $D_{\sigma+1}$;
- each positive occurrence of D in ϕ is interpreted by $D_{\tau+1}$;
- each (either negative or positive) occurrence of T in ϕ is interpreted by $T_{\xi+1}$.

We extend this notation to finite sets Γ of $\mathcal{L}$-sentences and write $\models \Gamma[\sigma, \tau, \xi]$ when $\models \phi[\sigma, \tau, \xi]$ for some $\phi \in \Gamma$. By Lemma 4.5.2, the sequence $\langle D_\xi \mid \xi \in On \rangle$ is a monotonically increasing, where On denotes the class of ordinals, and thus the next lemma can be shown in the standard manner.

LEMMA 6.6 (Persistence Lemma). *Let ϕ be an $\mathcal{L}$-sentences and Γ a finite set of $\mathcal{L}$-sentences. For every ordinals $\sigma_1 \leq \sigma_0$, $\tau_0 \leq \tau_1$, and ξ, the following hold.*

1. *If $\models \varphi[\sigma_0, \tau_0, \xi]$, then $\models \varphi[\sigma_1, \tau_1, \xi]$.*
2. *If $\models \Gamma[\sigma_0, \tau_0, \xi]$, then $\models \Gamma[\sigma_1, \tau_1, \xi]$.*

The next is the main lemma of the present section.

LEMMA 6.7. *For all ordinals δ, if $\mathsf{CD}^\infty \vdash^\delta_1 \Gamma$, then for all ordinals σ and ξ, if $\sigma + 2^\delta \leq \xi$, then $\models \Gamma[\sigma, \sigma + 2^\delta, \xi]$.*

PROOF. The claim is shown by induction on derivation. Throughout this proof, a, b, c, and d always denote closed $\mathcal{L}_0$-terms.

Suppose that the last inference is made by either of the following axioms:

$$\frac{}{\mathsf{CD}^\infty \vdash^\delta_1 \Gamma, A}\ (\mathsf{Ax1}) \qquad \frac{}{\mathsf{CD}^\infty \vdash^\delta_1 \Gamma, \mathrm{T}a, \neg\mathrm{T}b}\ (\mathsf{Ax3}) \qquad \frac{}{\mathsf{CD}^\infty \vdash^\delta_1 \Gamma, \mathrm{D}c}\ (\mathsf{D1}),$$

where A is a true closed $\mathcal{L}_0$-literal, $a^{\mathbb{N}} \approx^{\mathbb{N}} b^{\mathbb{N}}$, and $c^{\mathbb{N}}$ is the Gödel number of a closed $\mathcal{L}_0$-atomic. The claim for the first case is obvious; the claim for the second case follows from Proposition 4.4.2, and the claim for the third case follows from Proposition 4.2.1. Next, suppose the last inference is made by the axiom (Ax2):

$$\frac{}{\mathsf{CD}^\infty \vdash^\delta_1 \Gamma, \mathrm{D}a, \neg\mathrm{D}b}\ (\mathsf{Ax2}),$$

where $a^{\mathbb{N}} \approx^{\mathbb{N}} b^{\mathbb{N}}$. Take any ordinal σ and ξ. Since $D_{\sigma+1} \subset D_{\sigma+2^\delta+1}$ by Lemma 4.5.1, $a^{\mathbb{N}} \notin D_{\sigma+2^\delta+1}$ implies $a^{\mathbb{N}} \notin D_{\sigma+1}$ and thus $b^{\mathbb{N}} \notin D_{\sigma+1}$ by Proposition 4.4.1. Hence, either $a^{\mathbb{N}} \in D_{\sigma+2^\delta+1}$ or $b^{\mathbb{N}} \notin D_{\sigma+1}$ holds, and thus $\models \{\mathrm{D}a, \neg\mathrm{D}b\}[\sigma, \sigma + 2^\delta, \xi]$.

We move on to the induction step. The case where we use a logical rule in the last inference, except for cut, can be treated in the usual manner. For instance, suppose the last inference is made by $(\forall)$, namely, (our variant of) the ω-rule:

$$\frac{\mathsf{CD}^\infty \vdash^{\delta_a}_1 \Gamma, \forall x\, A(x), A(a), \text{ for each closed } \mathcal{L}_0\text{-term } a,}{\mathsf{CD}^\infty \vdash^\delta_1 \Gamma, \forall x\, A(x)}$$

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

where $\delta_a < \delta$ for each a. Take any σ and ξ with $\sigma + 2^\delta \leq \xi$. It follows from the induction hypothesis that for each closed term a, we have the following:

$$\models \Gamma \cup \{\forall x A(x)\} \cup \{A(a)\}\,[\sigma, \sigma + 2^{\delta_a}, \xi].$$

If $\models \Gamma \cup \{\forall x\, A(x)\}[\sigma, \sigma + 2^{\delta_a}, \xi]$ for any a, then we have $\models \Gamma \cup \{\forall x\, A(x)\}[\sigma, \sigma + 2^\delta, \xi]$ by Lemma 6.6. Otherwise, for all closed terms a, we have $\models A(a)[\sigma, \sigma + 2^{\delta_a}, \xi]$ and thus $\models A(a)[\sigma, \sigma + 2^\delta, \xi]$ by Lemma 6.6, which implies $\models \forall x\, A(x)[\sigma, \sigma + 2^\delta, \xi]$.

Next, the claim for the cases where the last inference is made by one of the rules for determinateness can be straightforwardly shown using Lemmata 4.2 and 4.5. For example, suppose the last inference is of the following form:

$$\frac{\mathsf{CD}^\infty \vdash_1^\gamma \Gamma, \neg \mathrm{D}a, (\neg \mathrm{D}b \vee \neg \mathrm{D}c) \wedge (\neg \mathrm{D}\dot{\neg} b \vee \neg \mathrm{F}b) \wedge (\neg \mathrm{D}\dot{\neg} c \vee \neg \mathrm{F}c)}{\mathsf{CD}^\infty \vdash_1^\delta \Gamma, \neg \mathrm{D}a}\ (\mathsf{D5}_\infty^{neg}),$$

where $\gamma < \delta$ and $a^{\mathbb{N}} = (b \dot{\wedge} c)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$. Take any σ and ξ with $\sigma + 2^\delta \leq \xi$. If we have $\models \Gamma \cup \{\neg \mathrm{D}a\}[\sigma, \sigma + 2^\gamma, \xi]$, then we get $\models \Gamma \cup \{\neg \mathrm{D}a\}[\sigma, \sigma + 2^\delta, \xi]$ by Lemma 6.6. Otherwise, we have the following by the induction hypothesis:

$$\models (\neg \mathrm{D}b \vee \neg \mathrm{D}c) \wedge (\neg \mathrm{D}\dot{\neg} b \vee \neg \mathrm{F}b) \wedge (\neg \mathrm{D}\dot{\neg} c \vee \neg \mathrm{F}c)\,[\sigma, \sigma + 2^\gamma, \xi].$$

Since $D_{\sigma+1} \supset D_\sigma$ and $D_\sigma \cap T_{\xi+1} = D_\sigma \cap T_\sigma$ by Lemma 4.5, this implies

$$b^{\mathbb{N}} \notin D_\sigma \text{ or } c^{\mathbb{N}} \notin D_\sigma, \text{ and } (\dot{\neg} b)^{\mathbb{N}} \notin D_\sigma \cap T_\sigma, \text{ and } (\dot{\neg} c)^{\mathbb{N}} \notin D_\sigma \cap T_\sigma.$$

Hence, by Lemma 4.2.5, we obtain $a^{\mathbb{N}} \notin D_{\sigma+1}$ and thus $\models \neg \mathrm{D}a[\sigma, \sigma + 2^\delta, \xi]$. Let us see one more example. Suppose the last inference is of the following form:

$$\frac{\mathsf{CD}^\infty \vdash_1^\gamma \Gamma, \mathrm{D}a, \forall s \mathrm{D}c(s/b) \vee \exists s (\mathrm{D}\dot{\neg} c(s/b) \wedge \mathrm{F}c(s/b))}{\mathsf{CD}^\infty \vdash_1^\delta \Gamma, \mathrm{D}a}\ (\mathsf{D6}_\infty^{pos}),$$

where $\gamma < \delta$ and $a^{\mathbb{N}} = (\dot{\forall} bc)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$. Take any σ and ξ with $\sigma + 2^\delta \leq \xi$. As before, if $\models \Gamma \cup \{\mathrm{D}a\}[\sigma, \sigma + 2^\gamma, \xi]$, then the claim follows by Lemma 6.6. Otherwise, by the induction hypothesis, we have

$$\models \forall s \mathrm{D}c(s/b) \vee \exists s (\mathrm{D}\dot{\neg} c(s/b) \wedge \mathrm{T}\dot{\neg} c(s/b))\,[\sigma, \sigma + 2^\gamma, \xi].$$

Since $\sigma + 2^\gamma + 1 \leq \sigma + 2^\delta \leq \xi$, it follows from Lemma 4.5 that

$$(c(d/b))^{\mathbb{N}} \in D_{\sigma+2^\delta} \text{ for all closed terms } d \text{ with } d^{\mathbb{N}} \in \mathrm{ClTerm}^{\mathbb{N}},$$
$$\text{or} (\dot{\neg} c(d/b))^{\mathbb{N}} \in D_{\sigma+2^\delta} \cap T_{\sigma+2^\delta} \text{ for some } d \text{ with } d^{\mathbb{N}} \in \mathrm{ClTerm}^{\mathbb{N}},$$

which implies $a^{\mathbb{N}} \in D_{\sigma+2^\delta+1}$ by Lemma 4.2.6 and thus $\models \mathrm{D}a[\sigma, \sigma + 2^\delta, \xi]$. We leave the other cases to the reader.

Now, let us assume that the last inference is made by a truth rule. Suppose the last inference is made by either $(\mathsf{T4}_\infty^{pos})$ or $(\mathsf{T3}_\infty^{neg})$:

$$\frac{\mathsf{CD}^\infty \vdash_1^\gamma \Gamma, \mathrm{T}a, \neg \mathrm{T}b}{\mathsf{CD}^\infty \vdash_1^\delta \Gamma, \mathrm{T}a}\ (\mathsf{T4}_\infty^{pos}) \qquad \frac{\mathsf{CD}^\infty \vdash_1^\gamma \Gamma, \neg \mathrm{T}a, \mathrm{T}b}{\mathsf{CD}^\infty \vdash_1^\delta \Gamma, \neg \mathrm{T}a}\ (\mathsf{T4}_\infty^{neg}),$$

where $\gamma < \delta$ and $a^{\mathbb{N}} = (\dot{\neg} b)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$. Take any σ and ξ with $\sigma + 2^\delta \leq \xi$. We will prove the claim only for the former case; the latter case can be similarly shown.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

If $\models \neg \mathrm{T}b[\sigma, \sigma + 2^\gamma, \xi]$, then $b^{\mathbb{N}} \notin T_{\xi+1}$ and thus $a^{\mathbb{N}} \in T_{\xi+1}$ by Lemma 4.3.4, which implies $\models \mathrm{T}a[\sigma, \sigma + 2^\delta, \xi]$. Otherwise, we have $\models \Gamma \cup \{\mathrm{T}a\}[\sigma, \sigma + 2^\gamma, \xi]$ by the induction hypothesis, and the claim follows by Lemma 6.6 as before. Next, suppose that we make either of the following inferences at the last step:

$$\frac{\mathsf{CD}^\infty \vdash_1^\gamma \Gamma, \mathrm{T}a, \mathrm{T}b^\circ \wedge \mathrm{D}b^\circ}{\mathsf{CD}^\infty \vdash_1^\delta \Gamma, \mathrm{T}a}\ (\mathsf{T3}_\infty^{pos}) \qquad \frac{\mathsf{CD}^\infty \vdash_1^\gamma \Gamma, \neg\mathrm{T}a, \neg\mathrm{T}b^\circ \wedge \mathrm{D}b^\circ}{\mathsf{CD}^\infty \vdash_1^\delta \Gamma, \neg\mathrm{T}a}\ (\mathsf{T3}_\infty^{neg}),$$

where $\gamma < \delta$ and $a^{\mathbb{N}} = (\dot{\mathrm{T}}b)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$. Take any σ and ξ with $\sigma + 2^\delta \leq \xi$. For the former case, if $\models \mathrm{T}b^\circ \wedge \mathrm{D}b^\circ[\sigma, \sigma + 2^\gamma, \xi]$, then, by Lemma 4.5, we have

$$(b^\circ)^{\mathbb{N}} \in D_{\sigma+2^\gamma+1} \cap T_{\xi+1} \subset D_{\sigma+2^\delta} \cap T_{\xi+1} = D_{\sigma+2^\delta} \cap T_{\sigma+2^\delta},$$

and thus, by Lemmata 4.2.2, 4.3.2, and 4.5, we obtain

$$a^{\mathbb{N}} \in D_{\sigma+2^\delta+1} \cap T_{\sigma+2^\delta+1} = D_{\sigma+2^\delta+1} \cap T_{\xi+1},$$

which implies $\models \mathrm{T}a[\sigma, \sigma + 2^\delta, \xi]$; otherwise, we have $\models \Gamma \cup \{\mathrm{T}a\}[\sigma, \sigma + 2^\gamma, \xi]$ by the induction hypothesis, and the claim follows by Lemma 6.6 as before. Similarly, for the latter case, if $\models \neg\mathrm{T}b^\circ \wedge \mathrm{D}b^\circ[\sigma, \sigma + 2^\gamma, \xi]$, then, by Lemma 4.5, we have

$$(b^\circ)^{\mathbb{N}} \in D_{\sigma+2^\gamma+1} \setminus T_{\xi+1} \subset D_{\sigma+2^\delta} \setminus T_{\xi+1} = D_{\sigma+2^\delta} \setminus T_{\sigma+2^\delta},$$

and thus, again by Lemmata 4.2.2, 4.3.2, and 4.5, we obtain

$$a^{\mathbb{N}} \in D_{\sigma+2^\delta+1} \setminus T_{\sigma+2^\delta+1} = D_{\sigma+2^\delta+1} \setminus T_{\xi+1},$$

which implies $\models \neg\mathrm{T}a[\sigma, \sigma + 2^\delta, \xi]$; otherwise, we have $\models \Gamma \cup \{\neg\mathrm{T}a\}[\sigma, \sigma + 2^\gamma, \xi]$ by the induction hypothesis, and the claim follows by Lemma 6.6 as before. Thirdly, suppose that the last inference is made by the rule $(\mathsf{T2}_\infty^{pos})$:

$$\frac{\mathsf{CD}^\infty \vdash_1^\gamma \Gamma, \mathrm{T}a, \mathrm{D}b^\circ}{\mathsf{CD}^\infty \vdash_1^\delta \Gamma, \mathrm{T}a}\ (\mathsf{T2}_\infty^{pos}),$$

where $\gamma < \delta$ and $a^{\mathbb{N}} = (\dot{\mathrm{D}}b)^{\mathbb{N}} \in \mathrm{Sent}^{\mathbb{N}}$. Take any σ and ξ with $\sigma + 2^\delta \leq \xi$. Suppose $\models \mathrm{D}b^\circ[\sigma, \sigma + 2^\gamma, \xi]$. We have $(b^\circ)^{\mathbb{N}} \in D_{\sigma+2^\gamma+1}$ and thus $(b^\circ)^{\mathbb{N}} \in D_\xi$ by Lemma 4.5.2. Hence, we get $a^{\mathbb{N}} \in T_{\xi+1}$ by Lemma 4.3.3 and thus $\models \mathrm{T}a[\sigma, \sigma + 2^\delta, \xi]$. Otherwise, the claim follows by Lemma 6.6 as before. The remaining cases for the other truth rules can be dealt with similarly.

Finally, let us assume that the last inference is made by (cut). We only consider the crucial case where the cut formulae are of the forms $\mathrm{D}a$ and $\neg\mathrm{D}a$:

$$\frac{\mathsf{CD}^\infty \vdash_1^{\gamma_0} \Gamma, \mathrm{D}a \qquad \mathsf{CD}^\infty \vdash_1^{\gamma_1} \Gamma, \neg\mathrm{D}a}{\mathsf{CD}^\infty \vdash_1^\delta \Gamma},$$

where $\gamma_0, \gamma_1 < \delta$. Let $\gamma = \max\{\gamma_0, \gamma_1\}$ and take any σ and ξ with $\sigma + 2^\delta \leq \xi$. Suppose $\not\models \Gamma[\sigma, \sigma + 2^\delta, \xi]$ for contradiction. Since $2^\gamma + 2^\gamma \leq 2^\delta$, it would follow that

$$\not\models \Gamma[\sigma, \sigma + 2^\gamma, \xi] \qquad \text{and} \qquad \not\models \Gamma[\sigma + 2^\gamma, \sigma + 2^\gamma + 2^\gamma, \xi]$$

by Lemma 6.6. Hence, by the induction hypothesis, we would have

$$\models \mathrm{D}a[\sigma, \sigma + 2^\gamma, \xi] \qquad \text{and} \qquad \models \neg\mathrm{D}a[\sigma + 2^\gamma, \sigma + 2^\gamma + 2^\gamma, \xi],$$

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

which means $a^{\mathbb{N}} \in D_{\sigma+2^{\gamma}+1}$ and $a^{\mathbb{N}} \notin D_{\sigma+2^{\gamma}+1}$; a contradiction. The remaining case in which the cut formulae are closed $\mathcal{L}_{\mathrm{T}}$-literals can be straightforwardly treated. ⊣

Combining this with Lemma 6.5, we obtain the next corollary.

COROLLARY 6.8. *For each $\mathcal{L}$-sentence ϕ, if $\mathsf{CD} \vdash \phi$, then there is some $\delta < \varepsilon_0$ such that $\models \psi[0, 2^{\delta}, 2^{\delta}]$.*

Now, the argument of this section can be adequately formalized within KF (or $\mathsf{RA}_{<\varepsilon_0}$), in which we can formalize the construction of D_{ξ}s and T_{ξ}s for sufficiently large ordinals ($< \varepsilon_0$) and prove (the arithmetization of) Corollary 6.8 in terms of them. Hence, in particular, we obtain the following.

COROLLARY 6.9. CD *is arithmetically conservative over* KF.

Combining this corollary with Corollary 5.6, we obtain our main theorem.

THEOREM 6.10. $\mathsf{CD}_0, \mathsf{CD}_1, \mathsf{CD}_2$, *and* CD *are arithmetically equivalent to* $\mathsf{KF}, \mathsf{RA}_{<\varepsilon_0}$, *and* $\mathsf{RT}_{<\varepsilon_0}$, *that is, they have exactly the same arithmetical theorems.*

In fact, our proofs establish stronger results: All the systems are proof-theoretically reducible to each other in the sense of Feferman [7, 9].

§7. Upper bound of the strength of CD^+. We will show that CD^+ is arithmetically conservative over $\mathsf{RT}_{<\varepsilon_{\varepsilon_0}}$. This will be shown by constructing a relative interpretation of CD^+ in $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!]$ $(= \mathsf{CT}[\![\mathsf{KF}]\!] + \mathrm{Cons})$ that preserves the arithmetical part.

7.1. Total and consistent predicates. We begin with the following definition:

$$D^+(x) \ :\Leftrightarrow\ (\mathrm{F}x \leftrightarrow \neg \mathrm{T}x).$$

Thus, $D^+(x)$ is equivalent to $(\mathrm{T}x \vee \mathrm{F}x) \wedge (\neg \mathrm{T}x \vee \neg \mathrm{F}x)$. We obviously have

$$\begin{aligned} &\mathsf{KF} + \mathrm{Cons} \vdash (\forall x \in \mathrm{Sent}_{\mathrm{T}})\big(D^+(x) \leftrightarrow (\mathrm{T}x \vee \mathrm{F}x)\big), \\ &\mathsf{KF} + \mathrm{Comp} \vdash (\forall x \in \mathrm{Sent}_{\mathrm{T}})\big(D^+(x) \leftrightarrow (\neg \mathrm{T}x \vee \neg \mathrm{F}x)\big). \end{aligned} \tag{11}$$

LEMMA 7.1. *The following are provable in* KF.

1. $\forall s\big(D^+(\dot{T}s) \leftrightarrow D^+(s^{\circ})\big)$; *use* $\mathsf{K2}$.
2. $(\forall x \in \mathrm{Sent}_{\mathrm{T}})\big(D^+(\dot{\neg}x) \leftrightarrow D^+x\big)$; *use* $\mathsf{K3}$.

LEMMA 7.2. *The following are provable in* $\mathsf{KF} + \mathrm{Cons}$.

1. $\forall x\, \forall y\Big(\mathrm{Sent}_{\mathrm{T}}(x \dot{\wedge} y) \rightarrow \Big(D^+(x \dot{\wedge} y) \leftrightarrow \big((D^+x \wedge D^+y) \vee (D^+x \wedge \mathrm{F}x) \vee (D^+y \wedge \mathrm{F}y)\big)\Big)\Big)$.
2. $\forall x\, \forall y\Big(\mathrm{Sent}_{\mathrm{T}}(x \dot{\vee} y) \rightarrow \Big(D^+(x \dot{\vee} y) \leftrightarrow \big((D^+x \wedge D^+y) \vee (D^+x \wedge \mathrm{T}x) \vee (D^+y \wedge \mathrm{T}y)\big)\Big)\Big)$.
3. $\forall v\, \forall x\Big(\mathrm{Sent}_{\mathrm{T}}(\dot{\forall} vx) \rightarrow \Big(D^+(\dot{\forall} vx) \leftrightarrow \Big(\forall s D^+x(s/v) \vee \exists s\big(D^+x(s/v) \wedge \mathrm{F}x(s/v)\big)\Big)\Big)\Big)$.
4. $\forall v\, \forall x\Big(\mathrm{Sent}_{\mathrm{T}}(\dot{\exists} vx) \rightarrow \Big(D^+(\dot{\exists} vx) \leftrightarrow \Big(\forall s D^+x(s/v) \vee \exists s\big(D^+x(s/v) \wedge \mathrm{T}x(s/v)\big)\Big)\Big)\Big)$.

PROOF. We work within $\mathsf{KF} + \mathrm{Cons}$; recall that we have $D^+(x) \leftrightarrow (\mathrm{T}x \vee \mathrm{F}x)$ for all $x \in \mathrm{Sent}_{\mathrm{T}}$ in $\mathsf{KF} + \mathrm{Cons}$. For claim 1, take any $x, y \in \mathrm{Sent}_{\mathrm{T}}$; then we have

$$\begin{aligned} \mathrm{T}(x \dot{\wedge} y) \vee \mathrm{F}(x \dot{\wedge} y) \quad &\overset{\mathsf{K4}\ \&\ \mathsf{K5}}{\Leftrightarrow} \quad (\mathrm{T}x \wedge \mathrm{T}y) \vee (\mathrm{F}x \vee \mathrm{F}y) \\ &\overset{\text{by logic}}{\Leftrightarrow} \quad (D^+x \wedge D^+y) \vee (D^+x \wedge \mathrm{F}x) \vee (D^+y \wedge \mathrm{F}y). \end{aligned}$$

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

Claim 2 can be shown similarly to 1 using Lemma 7.1.2 and K3. For claim 3, take any v and x with $\dot{\forall} vx \in \mathrm{Sent}_{\mathrm{T}}$; then we have

$$\begin{aligned} \mathrm{T}(\dot{\forall} vx) \vee \mathrm{F}(\dot{\forall} vx) &\overset{\mathsf{K6}\ \&\ \mathsf{K7}}{\Leftrightarrow} \forall s \mathrm{T}x(s/v) \vee \exists s \mathrm{F}x(a/v) \\ &\overset{\text{by logic}}{\Leftrightarrow} \forall s D^{+}x(s/v) \vee \exists s\big(D^{+}x(s/v) \wedge \mathrm{F}x(s/v)\big). \end{aligned}$$

Claim 4 can be shown similarly to 3 using Lemma 7.1.2 and K3. ⊣

LEMMA 7.3. $\mathsf{KF} \vdash \forall s\big(D^{+}(\dot{\mathrm{T}}s \vee \dot{\mathrm{F}}s) \leftrightarrow D^{+}s^{\circ}\big) \wedge \forall s\big(D^{+}(\dot{\neg}\dot{\mathrm{T}}s \vee \dot{\neg}\dot{\mathrm{F}}s) \leftrightarrow D^{+}s^{\circ}\big)$.

PROOF. Take any $s \in \mathrm{ClTerm}$. The following four equivalences are provable in KF:

$$\begin{array}{llll} \mathrm{T}(\dot{\mathrm{T}}s \vee \dot{\mathrm{F}}s) & \overset{5.2.1\ \&\ \mathsf{K2}}{\Leftrightarrow} & \mathrm{T}s^{\circ} \vee \mathrm{F}s^{\circ} & \overset{5.2.1\ \&\ \mathsf{K2}\ \&\ \mathsf{K3}}{\Leftrightarrow} \mathrm{T}(\dot{\neg}\dot{\mathrm{F}}s \vee \dot{\neg}\dot{\mathrm{T}}s); \\ \neg\mathrm{F}(\dot{\mathrm{T}}s \vee \dot{\mathrm{F}}s) & \overset{5.2.1\ \&\ \mathsf{K2}\ \&\ \mathsf{K3}}{\Leftrightarrow} & \neg\mathrm{F}s^{\circ} \vee \neg\mathrm{T}s^{\circ} & \overset{5.2.1\ \&\ \mathsf{K2}\ \&\ \mathsf{K3}}{\Leftrightarrow} \neg\mathrm{F}(\dot{\neg}\dot{\mathrm{F}}s \vee \dot{\neg}\dot{\mathrm{T}}s); \\ \mathrm{F}(\dot{\mathrm{T}}s \vee \dot{\mathrm{F}}s) & \overset{5.2.1\ \&\ \mathsf{K2}\ \&\ \mathsf{K3}}{\Leftrightarrow} & \mathrm{F}s^{\circ} \wedge \mathrm{T}s^{\circ} & \overset{5.2.1\ \&\ \mathsf{K2}\ \&\ \mathsf{K3}}{\Leftrightarrow} \mathrm{F}(\dot{\neg}\dot{\mathrm{F}}s \vee \dot{\neg}\dot{\mathrm{T}}s); \\ \neg\mathrm{T}(\dot{\mathrm{T}}s \vee \dot{\mathrm{F}}s) & \overset{5.2.1\ \&\ \mathsf{K2}}{\Leftrightarrow} & \neg\mathrm{T}s^{\circ} \wedge \neg\mathrm{F}s^{\circ} & \overset{5.2.1\ \&\ \mathsf{K2}\ \&\ \mathsf{K3}}{\Leftrightarrow} \neg\mathrm{T}(\dot{\neg}\dot{\mathrm{F}}s \vee \dot{\neg}\dot{\mathrm{T}}s). \end{array}$$

The claim follows from these by logic. ⊣

7.2. Reduction of CD^{+} to $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!]$. In the present subsection, we give an interpretation of CD^{+} in $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!]$ that preserves $\mathcal{L}_0$. For this purpose, we need a few preliminary definitions.

By the recursion theorem we obtain a recursive function f that satisfies the following condition:

$$f x = \begin{cases} x, & \text{if } x \in \mathrm{AtFml}_0, \\ \dot{\mathrm{T}}\dot{f}y, & \text{if } x = \dot{\mathrm{T}}y \text{ for } y \in \mathrm{Term}, \\ \dot{\mathrm{T}}\dot{f}y \vee \dot{\mathrm{F}}\dot{f}y, & \text{if } x = \dot{\mathrm{D}}y \text{ for } y \in \mathrm{Term}, \\ \dot{\neg}\dot{f}y, & \text{if } x = \dot{\neg}y \in \mathrm{Fml}, \\ (fy) \dot{\wedge} (fz), & \text{if } x = y \dot{\wedge} z \in \mathrm{Fml}, \\ \dot{\forall} v(fv), & \text{if } x = \dot{\forall} vy \in \mathrm{Fml}, \\ \ulcorner 0 \neq 0 \urcorner, & \text{otherwise.} \end{cases} \tag{12}$$

Here $\dot{f}$ is an arithmetical representation of the syntactic operation of applying the function f to terms so that $(\dot{f}s)^{\circ} = f(s^{\circ})$ is provable in PA for each $s \in \mathrm{ClTerm}$.[9] We can show that f is provably total in PA and thus f is a PA-definable function. Obviously we have $fx \in \mathrm{Fml}_{\mathrm{T}}$ for every $x \in \mathrm{Fml}$ (provably in PA), and, as in the case for the function k, we have $\forall z\big(\mathrm{Free}(z, x) \leftrightarrow \mathrm{Free}(z, fx)\big)$; hence, we have $fx \in \mathrm{Sent}_{\mathrm{T}}$ for every $x \in \mathrm{Sent}$ (provably in PA).

The next is shown by routine induction on the complexity of x (cf. Lemma 5.7).

LEMMA 7.4. $\mathsf{PA} \vdash (\forall x \in \mathrm{Fml})\forall s(\forall v \in \mathrm{Var})\big(f(x(s/v)) = (fx)(s/v)\big)$.

[9] The language $\mathcal{L}_0$ may not possess a function symbol for f depending on how one defines $\mathcal{L}_0$. In such a case, f is defined by some $\mathcal{L}_0$-formula $\phi(x, y)$, and $\dot{\mathrm{T}}\dot{f}z$ means $\ulcorner \forall w(\phi(x, w) \to \mathrm{T}w)\urcorner(z/\ulcorner x \urcorner)$. Then, we can easily verify that KF proves that $\mathrm{T}\ulcorner \forall w(\phi(x, w) \to \mathrm{T}w)\urcorner(z/\ulcorner x \urcorner)$ is equivalent to $\mathrm{T}f(z^{\circ})$.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

Now, we define a translation $\sharp$ of $\mathcal{L}$ into $\mathcal{L}_{\mathrm{T}}^{+}$ as follows: for each term a,

$$\mathrm{D}^{\sharp}a \;:=\; \mathrm{T}fa \vee \mathrm{F}fa \qquad \text{and} \qquad \mathrm{T}^{\sharp}a \;:=\; \mathbb{T}fa;$$

$\sharp$ preserves all the $\mathcal{L}_0$-vocabulary. By (11), we have

$$\mathsf{KF} + \mathrm{Cons} \vdash (\forall x \in \mathrm{Sent})\big(\mathrm{D}^{\sharp}x \leftrightarrow D^{+}(fx)\big). \tag{13}$$

We will show that $\sharp$ is a relative interpretation of CD in $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!]$.

Lemma 7.5. *The following are provable in* $\mathsf{KF} + \mathrm{Cons}$.

1. $\forall s \forall t \mathrm{D}^{\sharp}(s \doteq t)$; *by* (12) *and* K1.
2. $\forall s\big(\mathrm{D}^{\sharp}(\dot{T}s) \leftrightarrow \mathrm{D}^{\sharp}s^{\circ}\big)$; *by* (12), (13), *and Lemma* 7.1.1.
3. $(\forall x \in \mathrm{Sent})\big(\mathrm{D}^{\sharp}(\dot{\neg}x) \leftrightarrow \mathrm{D}^{\sharp}x\big)$; *by* (12), (13), *and Lemma* 7.1.2.

Hence, the $\sharp$*-translations of* D1, D2, *and* D4 *are provable in* $\mathsf{KF} + \mathrm{Cons}$.

Lemma 7.6. $\mathsf{KF} + \mathrm{Cons} \vdash \forall s\big(\mathrm{D}^{\sharp}(\dot{D}s) \leftrightarrow \mathrm{D}^{\sharp}s^{\circ}\big)$. *Hence, the* $\sharp$*-translations of* D3 *is provable in* $\mathsf{KF} + \mathrm{Cons}$.

Proof. The claim is readily observed from the following equivalences:

$$\mathrm{D}^{\sharp}(\dot{D}s) \overset{(13)}{\Leftrightarrow} D^{+}f(\dot{D}s) \overset{(12)}{\Leftrightarrow} D^{+}(\dot{T}fs \dot{\vee} \dot{F}fs) \overset{7.3}{\Leftrightarrow} D^{+}(fs)^{\circ} \Leftrightarrow D^{+}f(s^{\circ}) \overset{(13)}{\Leftrightarrow} \mathrm{D}^{\sharp}s^{\circ}.$$

⊣

Lemma 7.7. *The following are provable in* $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!]$.

1. $\forall s\, \forall t(\mathrm{T}^{\sharp}(s \doteq t) \leftrightarrow s^{\circ} = m^{\circ})$.
2. $\forall s\big(\mathrm{T}^{\sharp}(\dot{D}s) \leftrightarrow \mathrm{D}^{\sharp}s^{\circ}\big)$.
3. $(\forall x \in \mathrm{Sent})\big(\mathrm{T}^{\sharp}(\dot{\neg}x) \leftrightarrow \neg \mathrm{T}^{\sharp}x\big)\big)$.
4. $(\forall x, y \in \mathrm{Sent})\big(\mathrm{T}^{\sharp}(x \dot{\wedge} y) \leftrightarrow (\mathrm{T}^{\sharp}x \wedge \mathrm{T}^{\sharp}y)\big)$.
5. $\forall v \forall x\big(\forall \dot{v}x \in \mathrm{Sent} \rightarrow \big(\mathrm{T}^{\sharp}(\forall \dot{v}x) \leftrightarrow \forall s \mathrm{T}^{\sharp}x(s/v)\big)\big)$.

Hence, the $\sharp$*-translation of* T1 *and* T2 *and* T4–T6 *are provable in* $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!]$.

Proof. We work within $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!]$. Claim 1 is obvious by $\mathbb{T}1$. For claim 2, take any $s \in \mathrm{ClTerm}$, and then we infer

$$\mathbb{T}f(\dot{D}s) \overset{(12)}{\Leftrightarrow} \mathbb{T}(\dot{T}fs \dot{\vee} \dot{F}fs) \overset{5.4.1}{\Leftrightarrow} \mathbb{T}(\dot{T}fs) \vee \mathbb{T}(\dot{F}fs) \overset{\mathbb{T}1}{\Leftrightarrow} \mathrm{T}f(s^{\circ}) \vee \mathrm{F}f(s^{\circ}) \Leftrightarrow \mathrm{D}^{\sharp}s^{\circ}.$$

Claims 3 and 4 readily follow from $\mathbb{T}2$ and $\mathbb{T}3$, respectively, together with the definition (12) of f. For claim 5, let v and x be such that $\forall \dot{v}x \in \mathrm{Sent}$. Then, we have

$$\mathbb{T}f(\forall \dot{v}x) \overset{(12)}{\Leftrightarrow} \mathbb{T}\forall v f(x) \overset{\mathbb{T}4}{\Leftrightarrow} \forall s \mathbb{T}(fx)(s/v) \overset{7.4}{\Leftrightarrow} \forall s \mathbb{T}f(x(s/v)). \qquad \dashv$$

Lemma 7.8. $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!] \vdash \forall x\big(\mathrm{D}^{\sharp}(x) \rightarrow (\mathbb{T}f(x) \leftrightarrow \mathrm{T}f(x))\big)$.

Proof. If $x \notin \mathrm{Sent}$ then $f(x) = \ulcorner 0 \neq 0 \urcorner$ and thus $\mathbb{T}f(x) \wedge \mathrm{T}f(x)$. Let $x \in \mathrm{Sent}$. The claim is shown by induction on the complexity of x. For the base step, assume $x \in \mathrm{AtSent}$. The case where $x \in \mathrm{AtSent}_0$ is obvious. Next, if $x = \dot{T}s$ for some $s \in \mathrm{ClTerm}$, then we have

$$\mathbb{T}f(x) \overset{(12)}{\Leftrightarrow} \mathbb{T}(\dot{T}fs) \overset{\mathbb{T}1}{\Leftrightarrow} \mathrm{T}f(s^{\circ}) \overset{\mathsf{K}2}{\Leftrightarrow} \mathrm{T}\dot{T}fs \overset{(12)}{\Leftrightarrow} \mathrm{T}f(x).$$

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

Lastly, let $x = \dot{\mathrm{D}}s$ for some $s \in \mathrm{ClTerm}$. Then, we have

$$\mathbb{T}f(x) \overset{(12)\,\&}{\Leftrightarrow}{}^{5.4.1} \mathbb{T}(\dot{\mathrm{T}}fs) \vee \mathbb{T}(\dot{\mathrm{F}}fs) \overset{\mathbb{T}1}{\Leftrightarrow} \mathrm{T}fs^\circ \vee \mathrm{F}fs^\circ \overset{\mathsf{K2}}{\Leftrightarrow} \mathrm{T}(\dot{\mathrm{T}}fs) \vee \mathrm{T}(\dot{\mathrm{F}}fs) \overset{5.2.1\,\&\,(12)}{\Leftrightarrow} \mathrm{T}f(x).$$

For the induction step, we first let $x = \dot{\neg}y$ and assume $\mathrm{D}^\sharp x$. We have $\mathrm{D}^\sharp y$ by Lemma 7.5.3 and thus $D^+f(y)$ by (13). Hence, we have

$$\mathbb{T}f(x) \overset{(12)}{\Leftrightarrow} \mathbb{T}\dot{\neg}f(y) \overset{\mathbb{T}2}{\Leftrightarrow} \neg\mathbb{T}f(y) \overset{\mathrm{IH}}{\Leftrightarrow} \neg\mathrm{T}f(y) \overset{D^+f(y)}{\Leftrightarrow} \mathrm{F}f(y) \overset{(12)}{\Leftrightarrow} \mathrm{T}f(x);$$

note that the assumption $\mathrm{D}^\sharp x$ is crucial here. The other cases can be shown similarly, but without the need to use the assumption $\mathrm{D}^\sharp x$. ⊣

LEMMA 7.9. $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!] \vdash \forall s\big(\mathrm{D}^\sharp(s^\circ) \to (\mathrm{T}^\sharp(\dot{T}s) \leftrightarrow \mathrm{T}^\sharp(s^\circ))\big)$. *Hence, the* $\sharp$*-translation of* $\mathsf{T3}$ *is provable in* $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!]$.

PROOF. Take any $s \in \mathrm{ClTerm}$ and suppose $\mathrm{D}^\sharp(s^\circ)$. Then, we have

$$\mathbb{T}f(\dot{\mathrm{T}}s) \overset{(12)}{\Leftrightarrow} \mathbb{T}\dot{\mathrm{T}}fs \overset{\mathbb{T}1}{\Leftrightarrow} \mathrm{T}f(s^\circ) \overset{7.8}{\Leftrightarrow} \mathbb{T}f(s^\circ).$$ ⊣

LEMMA 7.10. $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!]$ *proves the following.*

1. $(\forall x, y \in \mathrm{Sent})\Big(\mathrm{D}^\sharp(x \dot{\wedge} y) \leftrightarrow \big((\mathrm{D}^\sharp x \wedge \mathrm{D}^\sharp y) \vee (\mathrm{D}^\sharp x \wedge \mathrm{F}^\sharp x) \vee (\mathrm{D}^\sharp y \wedge \mathrm{F}^\sharp y)\big)\Big)$.
2. $\forall v \forall x\Big(\dot{\forall} vx \in \mathrm{Sent} \to \Big(\mathrm{D}^\sharp(\dot{\forall} vx) \leftrightarrow \big(\forall s \mathrm{D}^\sharp x(s/v) \vee \exists s(\mathrm{D}^\sharp x(s/v) \wedge \mathrm{F}^\sharp x(s/v))\big)\Big)\Big)$.

Hence, the $\sharp$*-translations of* $\mathsf{D5}$ *and* $\mathsf{D6}$ *are provable in* $\mathsf{KF} + \mathrm{Cons}$.

PROOF. 1. Let $x, y \in \mathrm{Sent}$. Then, we have

$$\mathrm{D}^\sharp(x \dot{\wedge} y) \overset{(12)\,\&\,(13)}{\Leftrightarrow} D^+(fx \dot{\wedge} fy) \overset{7.2.1}{\Leftrightarrow} (D^+fx \wedge D^+fy) \vee (D^+fx \wedge \mathrm{F}fx) \vee (D^+fy \wedge \mathrm{F}fy) \overset{(12)\,\&\,(13)\,\&\,7.8}{\Leftrightarrow} (\mathrm{D}^\sharp x \wedge \mathrm{D}^\sharp y) \vee (\mathrm{D}^\sharp x \wedge \mathrm{F}^\sharp x) \vee (\mathrm{D}^\sharp y \wedge \mathrm{F}^\sharp y).$$

2. Let $\dot{\forall} vx \in \mathrm{Sent}$. Then, we similarly obtain

$$\mathrm{D}^\sharp(\dot{\forall} vx) \overset{(12)\,\&\,(13)\,\&\,7.2.3}{\Leftrightarrow} \Big(\forall s D^+(fx)(s/v) \vee \exists s\big(D^+(fx)(s/n) \wedge \mathrm{F}f(x)(s/v)\big)\Big) \overset{(12)\,\&\,(13)\,\&\,7.4\,\&\,7.8}{\Leftrightarrow} \Big(\forall s \mathrm{D}^\sharp x(s/v) \vee \exists s\big(\mathrm{D}^\sharp x(s/v) \wedge \mathrm{F}^\sharp x(s/v)\big)\Big).$$ ⊣

By Lemmata 7.5–7.7, 7.9, and 7.10, we obtain the next theorem.

THEOREM 7.11. *The translation* $\sharp$ *is a relative interpretation of* CD^+ *in* $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!]$. *Hence, in particular,* CD^+ *is arithmetically conservative over* $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!]$.

Combining this theorem with Corollary 5.10 and Fact 5.11, we obtain the next.

THEOREM 7.12. CD_2^+ *and* CD^+ *are arithmetically equivalent to* $\mathsf{RA}_{<\varepsilon_{\varepsilon_0}}$ *and* $\mathsf{RT}_{<\varepsilon_{\varepsilon_0}}$.

As before, our proofs actually establish that all these systems are proof-theoretically reducible to each other in the sense of Feferman [7, 9].

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

§8. Some supplementary results.

8.1. An interpretation of CD^+ in $\mathsf{CT}[\![\mathsf{KF}+\mathsf{Comp}]\!]$. Preliminarily, we give an interpretation of CD^+ in $\mathsf{CT}[\![\mathsf{KF}+\mathsf{Comp}]\!]$.[10]

Similarly to the function f, introduced in (12) above, we define a recursive function h by the recursion theorem so that the following condition obtains:

$$h(x) := \begin{cases} x, & \text{if } x \in \mathrm{AtFml}_0, \\ \dot{\mathrm{T}}\dot{h}y, & \text{if } x = \dot{\mathrm{T}}y \text{ for } y \in \mathrm{Term}, \\ \dot{\neg}\dot{\mathrm{T}}\dot{h}y \dot{\vee} \dot{\neg}\dot{\mathrm{F}}\dot{f}y, & \text{if } x = \dot{\mathrm{D}}y \text{ for } y \in \mathrm{Term}, \\ \dot{\neg}h(y), & \text{if } x = \dot{\neg}y \in \mathrm{Fml}, \\ h(y) \dot{\wedge} h(z), & \text{if } x = y \dot{\wedge} z \in \mathrm{Fml}, \\ \dot{\forall}v.h(y), & \text{if } x = \dot{\forall}vy \in \mathrm{Fml}, \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

Using h, the translation $\flat$ of $\mathcal{L}$ into $\mathcal{L}_\mathrm{T}$ is defined as follows: for each term a,

$$\mathrm{D}^\flat a \;=\; \neg \mathrm{T}ha \vee \neg \mathrm{F}ha \qquad \text{and} \qquad \mathrm{T}^\flat a \;=\; \mathbb{T}h(a);$$

all the $\mathcal{L}_0$-vocabulary, logical connectives, and quantifiers are preserved by $\flat$. By (11), we have $\mathsf{KF}+\mathsf{Comp} \vdash \forall x\big(\mathrm{D}^\flat x \leftrightarrow D^+(hx)\big)$.

The proof of the next theorem is completely dual to that of Theorem 7.11, and we omit the details.

THEOREM 8.1. *$\flat$ is a relative interpretation of* CD *in* $\mathsf{CT}[\![\mathsf{KF}+\mathsf{Comp}]\!]$.

8.2. Liars and truth tellers. In previous section we compared CD and its variants with other truth theories with respect to their strength. Truth theories are often compared by the way liar and truth teller sentences behave. In the present section we establish a few results about their behaviour in CD and its variants.

Let $\mathfrak{l}$ and $\mathfrak{t}$ be closed $\mathcal{L}_0$-terms with the following properties:

$$\mathsf{PA} \vdash \mathfrak{l} = \ulcorner \neg \mathrm{T}\mathfrak{l} \urcorner \quad \text{and} \quad \mathsf{PA} \vdash \mathfrak{t} = \ulcorner \mathrm{T}\mathfrak{t} \urcorner. \tag{15}$$

Hence, we have

$$\mathsf{CD} \vdash \mathrm{T}\mathfrak{l} \leftrightarrow \neg \mathrm{T}\ulcorner \mathrm{T}\mathfrak{l} \urcorner \quad \text{and} \quad \mathsf{CD} \vdash \mathrm{T}\mathfrak{t} \leftrightarrow \mathrm{T}\ulcorner \mathrm{T}\mathfrak{t} \urcorner \tag{16}$$

Therefore, $\neg \mathrm{T}\mathfrak{l}$ and $\mathrm{T}\mathfrak{t}$ are a liar and a truth-teller sentence, respectively.[11]

---

[10] Cantini's [5] 'dual' interpretation of $\mathsf{KF}+\mathsf{Cons}$ in $\mathsf{KF}+\mathsf{Comp}$ induces an interpretation of CD^+ in $\mathsf{CT}[\![\mathsf{KF}+\mathsf{Comp}]\!]$, but we will need a different interpretation for our purposes, namely, $\flat$ defined here.

[11] If the $\mathcal{L}_0$-vocabulary does not have $\mathfrak{l}$ and $\mathfrak{t}$ as genuine closed terms, they can still be taken as definable closed terms: in such a case, we take $\mathcal{L}_0$-formulae $\phi(x)$ and $\psi(x)$ with exactly one free variable such that

$$\mathsf{PA} \vdash \forall x\big(\phi(x) \to x = \ulcorner \forall x(\phi(x) \to \neg \mathrm{T}x) \urcorner\big) \quad \text{and} \quad \mathsf{PA} \vdash \forall x\big(\psi(x) \to x = \ulcorner \forall x(\psi(x) \to \mathrm{T}x) \urcorner\big),$$

and (16) means that

$$\mathsf{CD} \vdash \forall x\big(\phi(x) \to (\mathrm{T}x \leftrightarrow \neg \mathrm{T}\ulcorner \forall x(\phi(x) \to \neg \mathrm{T}x) \urcorner)\big) \quad \text{and} \quad \mathsf{CD} \vdash \forall x\big(\psi(x) \to (\mathrm{T}x \leftrightarrow \neg \mathrm{T}\ulcorner \forall x(\psi(x) \to \mathrm{T}x) \urcorner)\big),$$

both of which are readily verified. All the results in this section are true in terms of these paraphrases.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

PROPOSITION 8.2.

1. $\mathsf{CD} \vdash \neg \mathrm{D}\mathfrak{l}$.
2. $\mathsf{CD}^+ \nvdash \mathrm{D}\mathfrak{t}$.
3. $\mathsf{CD}^+ \nvdash \mathrm{T}\mathfrak{t}$ *and* $\mathsf{CD}^+ \nvdash \neg\mathrm{T}\mathfrak{t}$.
4. $\mathsf{CD}^+ \nvdash \neg\mathrm{T}\mathfrak{l}$ *and* $\mathsf{CD}^+ \nvdash \mathrm{T}\mathfrak{l}$.

PROOF. 1. Provably in CD, if $\mathrm{D}\mathfrak{l}$ were the case, then we would have $\mathrm{T}\mathfrak{l} \leftrightarrow \mathrm{T}\ulcorner\mathrm{T}\mathfrak{l}\urcorner$ by T3, which contradicts (16).

2. Using (15), we can show that $\mathfrak{t} \notin D_\xi$ for all ξ by straightforward induction on ξ.

3. For the function f defined in Section 7.2, we observe that PA proves the following:

$$f(\mathfrak{t}) = f(\ulcorner\mathrm{T}\mathfrak{t}\urcorner) = \ulcorner\mathrm{T}f(\mathfrak{t})\urcorner \quad \text{and} \quad h(\mathfrak{t}) = h(\ulcorner\mathrm{T}\mathfrak{t}\urcorner) = \ulcorner\mathrm{T}h(\mathfrak{t})\urcorner; \tag{17}$$

hence, both $f(\mathfrak{t})$ and $h(\mathfrak{t})$ are also truth-tellers. Let I be the least Kripkean fixed point, namely, the least set $I \subset \omega$ such that $(\mathbb{N}, I) \models \mathsf{KF}$, where I interprets the truth predicate T. It is known that $(\mathbb{N}, I) \models \mathrm{Cons}$. By the standard argument, we can show that $f(\mathfrak{t}) \notin I$ and $\dot{\neg}h(\mathfrak{t}) \notin I$. Let $J := \{\#\neg\phi \mid \#\phi \notin I\}$; hence, $(\mathbb{N}, J)$ is the $\mathcal{L}_\mathrm{T}$-structure induced from $(\mathbb{N}, I)$ by Cantini's dual interpretation, and thus $(\mathbb{N}, J) \models \mathsf{KF} + \mathrm{Comp}$. Then, we have

$$(\mathbb{N}, I) \not\models \mathrm{T}f(\mathfrak{t}) \quad \text{and} \quad (\mathbb{N}, J) \models \mathrm{T}h(\mathfrak{t}). \tag{18}$$

Now, by (17), we have the following:

$$\mathsf{CT}[\![\mathsf{KF}]\!] \vdash \mathbb{T}f(\mathfrak{t}) \leftrightarrow \mathrm{T}f(\mathfrak{t}) \quad \text{and} \quad \mathsf{CT}[\![\mathsf{KF}]\!] \vdash \mathbb{T}h(\mathfrak{t}) \leftrightarrow \mathrm{T}h(\mathfrak{t}). \tag{19}$$

Let $I^+ := \{\#\phi \mid (\mathbb{N}, I) \models \phi\}$ and $J^+ := \{\#\phi \mid (\mathbb{N}, J) \models \phi\}$; then, the $\mathcal{L}_\mathrm{T}^+$-structures $(\mathbb{N}, I, I^+)$ and $(\mathbb{N}, J, J^+)$, in which $\mathbb{T}$ is interpreted by I^+ and J^+ respectively, are models of $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!]$ and $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Comp}]\!]$, respectively. It follows from (18) and (19) that

$$(\mathbb{N}, I, I^+) \not\models \mathbb{T}f(\mathfrak{t}) \quad \text{and} \quad (\mathbb{N}, J, J^+) \not\models \neg\mathbb{T}h(\mathfrak{t}).$$

Hence, $\mathsf{CD}^+ \nvdash \mathrm{T}\mathfrak{t}$ and $\mathsf{CD}^+ \nvdash \neg\mathrm{T}\mathfrak{t}$ by Theorems 7.11 and 8.1.

4. Observe that PA proves the following:

$$f(\mathfrak{l}) = f(\ulcorner\neg\mathrm{T}\mathfrak{l}\urcorner) = \ulcorner\neg\mathrm{T}f(\mathfrak{l})\urcorner \quad \text{and} \quad h(\mathfrak{l}) = h(\ulcorner\neg\mathrm{T}\mathfrak{l}\urcorner) = \ulcorner\neg\mathrm{T}h(\mathfrak{l})\urcorner; \tag{20}$$

namely, $f(\mathfrak{l})$ and $h(\mathfrak{l})$ are also liar sentences. Hence, we obtain

$$\mathsf{KF} \vdash \mathrm{T}f(\mathfrak{l}) \leftrightarrow \mathrm{F}f(\mathfrak{l}) \qquad \text{and} \qquad \mathsf{KF} \vdash \mathrm{T}h(\mathfrak{l}) \leftrightarrow \mathrm{F}h(\mathfrak{l}), \tag{21}$$

which implies $\mathsf{KF} + \mathrm{Cons} \vdash \neg\mathrm{T}f(\mathfrak{l})$ and $\mathsf{KF} + \mathrm{Comp} \vdash \mathrm{T}h(\mathfrak{l})$. Finally, it follows from (20) that $\mathsf{CT}[\![\mathsf{KF}]\!]$ proves

$$\mathbb{T}f(\mathfrak{l}) \leftrightarrow \mathbb{T}\ulcorner\neg\mathrm{T}f(\mathfrak{l})\urcorner \leftrightarrow \neg\mathrm{T}f(\mathfrak{l}) \quad \text{and} \quad \mathbb{T}h(\mathfrak{l}) \leftrightarrow \mathbb{T}(\ulcorner\neg\mathrm{T}h\mathfrak{l}\urcorner) \;\Leftrightarrow\; \neg\mathrm{T}h(\mathfrak{l}).$$

Hence, we have $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!] \vdash \mathbb{T}f(\mathfrak{l})$ and $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Comp}]\!] \vdash \neg\mathbb{T}h(\mathfrak{l})$, which implies $\mathsf{CD}^+ \nvdash \neg\mathrm{T}\mathfrak{l}$ and $\mathsf{CD}^+ \nvdash \mathrm{T}\mathfrak{l}$ by Theorems 7.11 and 8.1. ⊣

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

PROPOSITION 8.3. *Let us write λ for $\neg\mathrm{T}\iota$. We have the following*:

1. $\mathsf{CD}^+ \nvdash \mathrm{T}\ulcorner\mathrm{T}\ulcorner\lambda \vee \neg\lambda\urcorner\urcorner$ *and* $\mathsf{CD}^+ \nvdash \neg\mathrm{T}\ulcorner\mathrm{T}\ulcorner\lambda \vee \neg\lambda\urcorner\urcorner$; *hence, the truth of some logical truth is neither provable nor refutable in* CD.
2. $\mathsf{CD}^+ \nvdash \mathrm{T}\ulcorner(\mathrm{T}\ulcorner\neg\lambda\urcorner \leftrightarrow \neg\mathrm{T}\ulcorner\lambda\urcorner)\urcorner$; *hence, not all* CD*-axioms are provably true in* CD.
3. $\mathsf{CD}^+ \vdash \neg\mathrm{D}\ulcorner(\mathrm{T}\ulcorner\neg\lambda\urcorner \leftrightarrow \neg\mathrm{T}\ulcorner\lambda\urcorner)\urcorner$; *hence, not all* CD*-axioms are provably determinate in* CD.

PROOF. 1. $\mathsf{CT}[\![\mathsf{KF}]\!]$ proves the following equivalences:

$$\mathrm{T}^\sharp\ulcorner\mathrm{T}\ulcorner\lambda \vee \neg\lambda\urcorner\urcorner \Leftrightarrow \mathbb{T}\ulcorner\mathrm{T}f(\ulcorner\lambda \vee \neg\lambda\urcorner)\urcorner \Leftrightarrow \mathrm{T}f(\ulcorner\lambda \vee \neg\lambda\urcorner) \Leftrightarrow \mathrm{T}f(\lambda) \vee \mathrm{F}f(\lambda),$$

$$\mathrm{T}^\flat\ulcorner\mathrm{T}\ulcorner\lambda \vee \neg\lambda\urcorner\urcorner \Leftrightarrow \mathbb{T}\ulcorner\mathrm{T}h(\ulcorner\lambda \vee \neg\lambda\urcorner)\urcorner \Leftrightarrow \mathrm{T}h(\ulcorner\lambda \vee \neg\lambda\urcorner) \Leftrightarrow \mathrm{T}h(\lambda) \vee \mathrm{F}h(\lambda).$$

By (15) and (21), $\mathsf{KF} + \mathrm{Cons} \vdash \neg(\mathrm{T}f(\lambda) \vee \mathrm{F}f(\lambda))$ and $\mathsf{KF} + \mathrm{Comp} \vdash \mathrm{T}h(\lambda) \vee \mathrm{F}h(\lambda)$. Hence, $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!] \vdash \neg\mathrm{T}^\sharp\ulcorner\mathrm{T}\ulcorner\lambda \vee \neg\lambda\urcorner\urcorner$ and $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Comp}]\!] \vdash \mathrm{T}^\flat\ulcorner\mathrm{T}\ulcorner\lambda \vee \neg\lambda\urcorner\urcorner$; thus, $\mathsf{CD}^+ \nvdash \mathrm{T}\ulcorner\mathrm{T}\ulcorner\lambda \vee \neg\lambda\urcorner\urcorner$ and $\mathsf{CD}^+ \nvdash \neg\mathrm{T}\ulcorner\mathrm{T}\ulcorner\lambda \vee \neg\lambda\urcorner\urcorner$ by Theorems 7.11 and 8.1.

2. $\mathsf{CT}[\![\mathsf{KF}]\!]$ derives the following:

$$\begin{aligned}\mathrm{T}^\sharp\ulcorner(\mathrm{T}\ulcorner\neg\lambda\urcorner \leftrightarrow \neg\mathrm{T}\ulcorner\lambda\urcorner)\urcorner \quad &\Leftrightarrow \quad \mathbb{T}f(\ulcorner\mathrm{T}\ulcorner\neg\lambda\urcorner\urcorner) \leftrightarrow \neg\mathbb{T}f(\ulcorner\mathrm{T}\ulcorner\lambda\urcorner\urcorner)\\ &\Leftrightarrow \quad \mathrm{F}f(\lambda) \leftrightarrow \neg\mathrm{T}f(\lambda)\\ &\overset{(15)\,\&\,(21)}{\Leftrightarrow} \mathrm{T}f(\lambda) \leftrightarrow \neg\mathrm{T}f(\lambda).\end{aligned}$$

Hence, we have $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!] \vdash \neg\mathrm{T}^\sharp\ulcorner(\mathrm{T}\ulcorner\neg\lambda\urcorner \leftrightarrow \neg\mathrm{T}\ulcorner\lambda\urcorner)\urcorner$, from which the claim follows by Theorems 7.11.

3. Provably in CD, if $\mathrm{D}\ulcorner(\mathrm{T}\ulcorner\neg\lambda\urcorner \leftrightarrow \neg\mathrm{T}\ulcorner\lambda\urcorner)\urcorner$, then $\mathrm{D}\ulcorner\neg\mathrm{T}\ulcorner\lambda\urcorner\urcorner$ and thus $\mathrm{D}\iota$ by (15), which contradicts Proposition 8.2.1. ⊣

8.3. Additional axioms. In this subsection, we will consider a few conservative extensions of CD^+ with additional axioms about iterations of truth.

LEMMA 8.4.

1. $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!] \vdash \forall s(\mathrm{T}^\sharp\dot{\mathrm{T}}s \to \mathrm{T}^\sharp s^\circ)$.
2. $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!] \vdash \forall s(\mathrm{F}^\sharp s^\circ \to \mathrm{F}^\sharp\dot{\mathrm{T}}s)$.

PROOF. We work within $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Cons}]\!]$. Take any $s \in \mathrm{ClTerm}$. We have

$$\mathbb{T}f(\dot{\mathrm{T}}s) \overset{(12)\,\&\,\mathbb{T}1}{\Leftrightarrow} \mathrm{T}f(s^\circ) \overset{\text{by logic}}{\Leftrightarrow} D^+f(s^\circ) \wedge \mathrm{T}f(s^\circ) \overset{7.8}{\Rightarrow} \mathbb{T}f(s^\circ);$$

recall that $\mathrm{T}^\sharp a := \mathbb{T}fa$ and $\mathrm{D}^\sharp a := D^+fa$. Similarly, for each $s \in \mathrm{ClTerm}$, we have

$$\neg\mathbb{T}f(\dot{\neg}\dot{\mathrm{T}}s) \overset{(12)\,\&\,\mathbb{T}2}{\Leftrightarrow} \mathbb{T}\dot{\mathrm{T}}fs \overset{\mathbb{T}1}{\Leftrightarrow} \mathrm{T}f(s^\circ) \overset{\text{by logic}}{\Leftrightarrow} D^+f(s^\circ) \wedge \mathrm{T}f(s^\circ) \overset{7.8}{\Rightarrow} \mathbb{T}f(s^\circ) \overset{(12)\,\&\,\mathbb{T}2}{\Leftrightarrow} \neg\mathbb{T}f(\dot{\neg}s^\circ).$$

This completes the proof. ⊣

By the dual argument, we can also show the following.

LEMMA 8.5.

1. $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Comp}]\!] \vdash \forall s(\mathrm{F}^\flat\dot{\mathrm{T}}s \to \mathrm{F}^\flat s^\circ)$.
2. $\mathsf{CT}[\![\mathsf{KF} + \mathrm{Comp}]\!] \vdash \forall s(\mathrm{T}^\flat s^\circ \to \mathrm{T}^\flat\dot{\mathrm{T}}s)$.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

Let $\mathsf{num}(n)$ represent the function $n \mapsto \#\overline{n}$ and thus it is the code of the numeral for n. Following the notation of [21], given a formula $\phi(x)$ with a distinguished free variable x, $\ulcorner\phi(\dot{z})\urcorner$ denotes $\ulcorner\phi(x)\urcorner(\mathsf{num}(z)/\ulcorner x\urcorner)$ and also $\ulcorner\phi(\dot{s})\urcorner$ denotes $\ulcorner\phi(x)\urcorner(s/\ulcorner x\urcorner)$ for $s \in \mathrm{ClTerm}$. We will occasionally write $\ulcorner\phi(\dot{z})\urcorner$ or $\ulcorner\phi(\dot{s})\urcorner$ without specifying the distinguished free variable of ϕ, but in doing so we always assume that ϕ contains some free variable to be substituted for.

LEMMA 8.6. $\mathsf{CT}[\![\mathsf{KF}]\!] \vdash \forall s(\mathrm{T}^\sharp\ulcorner\mathrm{T}\dot{\mathrm{T}}\dot{s}\urcorner \leftrightarrow \mathrm{T}^\sharp\dot{\mathrm{T}}s) \wedge \forall s(\mathrm{T}^\flat\ulcorner\mathrm{T}\dot{\mathrm{T}}\dot{s}\urcorner \leftrightarrow \mathrm{T}^\flat\dot{\mathrm{T}}s)$.

PROOF. Take any $s \in \mathrm{ClTerm}$. The first conjunct is shown as follows:

$$\mathbb{T}f(\ulcorner\mathrm{T}\dot{\mathrm{T}}\dot{s}\urcorner) \Leftrightarrow \mathbb{T}\ulcorner\mathrm{T}f(\dot{\mathrm{T}}\dot{s})\urcorner \Leftrightarrow \mathrm{T}f(\dot{\mathrm{T}}s) \Leftrightarrow \mathrm{T}\dot{\mathrm{T}}fs \Leftrightarrow \mathrm{T}fs^\circ \Leftrightarrow \mathbb{T}f(\dot{\mathrm{T}}s).$$

The second conjunct can be shown similarly by just replacing f above with h. ⊣

We conclude with the next theorem, which immediately follows from the last three lemmata and Theorems 7.11 and 8.1.

THEOREM 8.7. *The following two systems are both conservative over* CD^+:

1. $\mathsf{CD}^+ + \forall s(\mathrm{T}\dot{\mathrm{T}}s \to \mathrm{T}s^\circ) + \forall s(\mathrm{F}s^\circ \to \mathrm{F}\dot{\mathrm{T}}s) + \forall s(\mathrm{T}\ulcorner\mathrm{T}\dot{\mathrm{T}}\dot{s}\urcorner \leftrightarrow \mathrm{T}\dot{\mathrm{T}}s)$.
2. $\mathsf{CD}^+ + \forall s(\mathrm{F}\dot{\mathrm{T}}s \to \mathrm{F}s^\circ) + \forall s(\mathrm{T}s^\circ \to \mathrm{T}\dot{\mathrm{T}}s) + \forall s(\mathrm{T}\ulcorner\mathrm{T}\dot{\mathrm{T}}\dot{s}\urcorner \leftrightarrow \mathrm{T}\dot{\mathrm{T}}s)$.

Hence, either set of the new axioms can be consistently added to CD.

§9. The minimality axiom. The axioms for determinateness postulate that certain sentences are determinate; but CD lacks axioms that tell us that no other sentences are determinate. For instance, our axioms for determinateness do not rule out that truth teller sentences or similar sentences are determinate. Only in the case of sentences, such as liar sentences, where the assumption of determinateness leads to inconsistency, can we prove in CD that they are not determinate.

Theorem 4.13 says that no matter what model $(\mathbb{N}, X, Y)$ of CD we choose, D_∞ is the least fixed point of $\Gamma_{\mathcal{D}[Y]}$ and can be inductively defined 'from the bottom up' *in* Y. Hence, D_∞ seems to be a natural candidate for the 'intended' interpretation of the determinateness predicate D. In this section, we present an axiomatization of this conception of determinateness, which allows us to refute the determinateness of truth tellers, and state the proof-theoretic strength of the resulting system; however, the full analysis of the system goes beyond limited space of the present paper and is left for the Part II.

The system ID_1 of (positive arithmetical) inductive definitions contains an axiom schema expressing that each fixed point expressed by a predicate of ID_1 is the least, inductively defined, one. We use the same strategy for expressing the leastness of D. We remind the reader of the definition of ID_1. The language $\mathcal{L}_{\mathrm{fix}}$ of ID_1 has a predicate symbol $J_\mathcal{A}$ for each second-order arithmetical formula $\mathcal{A}(x, X)$ with only the displayed variables free in which X occurs only positively. ID_1 comprises the following axioms for each predicate $J_\mathcal{A}$ together with the axioms of PA and full induction for $\mathcal{L}_{\mathrm{fix}}$:

ID1 $\forall x\big(\mathcal{A}(x, J_\mathcal{A}) \to J_\mathcal{A}x\big)$,
ID2 $\forall x\big(\mathcal{A}(x, \Phi) \to \Phi(x)\big) \to \forall x\big(J_\mathcal{A}x \to \Phi(x)\big)$, for all $\Phi \in \mathcal{L}_{\mathrm{fix}}$,

where $\mathcal{A}(x, \Phi)$ for an $\mathcal{L}_{\mathrm{fix}}$-formula $\Phi(u)$ with a designated variable u is the result of substituting $\Phi(a)$ for each occurrence of $a \in X$ in $\mathcal{A}(x, X)$ for each term a

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

(with renaming of bound variables as necessary to avoid collision). Inspired by ID2, Burgess [4] proposes the following axiom schema as an axiomatic characterization of the *groundedness* of the truth predicate T in Kripke's [28] sense:

$$\forall x\big(\mathcal{B}(x,\Phi)\to\Phi(x)\big)\to\forall x\big(\mathrm{T}x\to\Phi(x)\big),\ \text{for all } \Phi\in\mathcal{L}_{\mathrm{T}},$$

where $\mathcal{B}(x,X)$ is taken so that $\forall x\big(\mathcal{B}(x,X)\to X\big)$ expresses that X is closed under the strong Kleene evaluation schema of truth values. Burgess's schema expresses that the extension of T is the least Kripkean fixed point under the strong Kleene evaluation schema; the system obtained by augmenting KF with this schema is called KFB. We follow the lead of Burgess's idea toward an axiomatic characterization of D as the least set that satisfies the closure conditions of determinateness relative to the truth predicate T (but recall that the least such set is invariant across the choice of T).

For each $\mathcal{L}$-formula $\Phi(u)$ with a designated variable u, we write $\mathcal{D}(x,\Phi)$ for the result of substituting $\Phi(a)$ for each occurrence of $\mathrm{D}a$ in $\mathcal{D}(x)$ for each term a (with renaming of bound variables as necessary to avoid collision). For the uniformity of notation, we will write $\mathcal{D}(x,\mathrm{D})$ for $\mathcal{D}(x)$ in (and only in) the present section. Thereby we introduce two new axioms.

$D_\mu 1$ $\quad\forall x(\mathcal{D}(x,\mathrm{D})\to\mathrm{D}x).$

$D_\mu 2$ $\quad\forall x\big(\mathcal{D}(x,\Phi)\to\Phi(x)\big)\to\forall x(\mathrm{D}x\to\Phi(x)),$ for all $\Phi\in\mathcal{L}$.

Note that $\mathrm{D}_\mu 2$ is not a single axiom but an axioms schema.

Definition 9.1. We define the system CD_μ as $\mathsf{CD}+\mathrm{D}_\mu 1+\mathrm{D}_\mu 2$. We can easily show that $\mathrm{D}\mu 1$ is derivable from CD and thus CD_μ is identical with $\mathsf{CD}+\mathrm{D}_\mu 2$. The system CD^+_μ is defined as $\mathsf{CD}^+ + \mathrm{D}_\mu 2$.

The next follows from Theorems 4.8 and 4.13.

Corollary 9.2. $(\mathbb{N}, D_\infty, \mathbb{T})\models\mathsf{CD}^+_\mu$.

The system CD_μ has some desirable properties. For example, we have the following.

Proposition 9.3. $\mathsf{CD}_\mu\vdash\neg\mathrm{D}\mathfrak{t}$; *see Section 8.2 for the definition of* $\mathfrak{t}$.

Proof. Suppose $\mathrm{D}\mathfrak{t}$ for contradiction. Let $\Phi(x):=\mathrm{D}x\wedge x\neq\mathfrak{t}$. Since $\mathcal{D}(\mathfrak{t},\Phi)$ implies $\Phi(\mathfrak{t})$ by (15), we have $\neg\mathcal{D}(\mathfrak{t},\Phi)$ and thus $\forall x(\mathcal{D}(x,\Phi)\to\Phi(x))$ by $\mathrm{D}_\mu 1$, which would imply $\forall x(\mathrm{D}x\to\Phi(x))$; a contradiction. ⊣

In fact, all the underivability results proved in Section 8.2 still hold in CD_μ and CD^+_μ, and the additional axioms considered in Section 8.3 are also consistent with CD_μ and CD^+_μ.

The full proof-theoretic analyses of these new systems are left for the Part II. We state the main results below without giving their proofs.

Theorem 9.4.

1. CD_μ *is proof-theoretically equivalent to* ID_1.
2. CD^+_μ *is proof-theoretically equivalent to* BID^2_1 (*see* [36] *for its definition and Fujimoto* [15] *for its proof-theoretic analysis*), *and thus its proof-theoretic ordinal* (*if appropriately defined*) *is* $\psi_\Omega(\varepsilon_{\varepsilon_0})$.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

§10. Comparison with other axiomatizations of truth. We think that CD is more promising than many other systems found in the literature. This section contains brief comparisons with such systems. We leave a more thorough discussion of the philosophical aspects for another paper.

10.1. Comparison with the Friedman–Sheard system. The Friedman–Sheard system FS features all the truth axioms of CD, except for the truth iteration axiom T3 and all axioms involving D, which is not in the language of FS. Axiom T3 is replaced with two rules called Necessitation NEC and Conecessitation CONEC:

$$\frac{\phi}{\mathrm{T}\ulcorner\phi\urcorner}\ \mathsf{NEC} \qquad \frac{\mathrm{T}\ulcorner\phi\urcorner}{\phi}\ \mathsf{CONEC}.$$

Adding these two rules for all sentences to T1 and T4 – T6 yields the Friedman–Sheard system FS, which was analyzed by Friedman and Sheard [13] with a different axiomatization. McGee [31] showed that a subsystem of FS is ω-inconsistent and thus does not have a standard model, that is, the standard model of arithmetic cannot be expanded to a model of FS by specifying a suitable extension for the truth predicate. Halbach [18] determined the proof-theoretic strength of FS as ramified analysis $\mathsf{RA}_{<\omega}$ up to ω or ω-times iterated typed truth (see Halbach [21, Section 14.2]).

A neat feature of FS is its symmetry: What is provable in FS and what is provably true in FS coincide. Moreover, if ϕ is a classical tautology such as $\lambda \vee \neg\lambda$ (for a liar sentence λ), the sentence

$$\underbrace{\mathrm{T}\ulcorner\mathrm{T}\ulcorner\ldots\mathrm{T}}_{n \text{ iterations of T}}\ \ulcorner\phi\urcorner\ldots\urcorner\urcorner$$

is provable in FS. In CD $\mathrm{T}\ulcorner\phi\urcorner$ is provable for every classical tautology, but $\mathrm{T}\ulcorner\mathrm{T}\ulcorner\lambda \vee \neg\lambda\urcorner\urcorner$ and further iterations are not, as was shown in Proposition 8.3. In FS no transfinite iterations of truth can be proved. Hence neither is FS a subsystem of CD nor is CD a subsystem of FS. Moreover, neither can the truth predicate of FS be defined in CD nor can the truth predicate of CD be defined in FS. This follows from general arguments by Fujimoto [14] (see also Halbach [21]). We will show in Part II that CD can be consistently closed under NEC and CONEC. The result is a system that properly contains FS and is consequently ω-inconsistent.

The system FS has rightly been criticized by various authors. Barrio and Picollo [2] list some of the problematic features of FS caused by the ω-inconsistency. In particular, the rule NEC can naturally be replaced with an ω-times iterated reflection axiom. If the reflection axiom is iterated into the transfinite, an inconsistency ensues (Halbach and Horsten [3] and Halbach [21, Corollary 14.39]). These shortcomings of FS prompted our search for an alternative to FS without giving up the thorough classicality of FS and endorsing the internal non-classicality of systems such as KF.

Some alternative systems have been developed that prove the truth of classical tautologies, but are ω-consistent such as Cantini's [6] VF and Stern's [41] IT and their variants. For these systems ω-models can be obtained via constructions involving supervaluations.

10.2. Comparison with Kripke–Feferman. Feferman [8] defined a determinateness predicate D in terms of a truth and a falsity predicate within the Kripke–Feferman

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

system KF.[12] Understanding the falsity of a sentence as the truth of its negation, Feferman defined D as the formula we called D^+ above:

$$Dx \leftrightarrow \mathrm{Sent}(x) \wedge (\mathrm{T}x \vee \mathrm{F}x) \wedge \neg(\mathrm{T}x \wedge \mathrm{F}x).$$

The schema $D\ulcorner\phi\urcorner \rightarrow (\mathrm{T}\ulcorner\phi\urcorner \leftrightarrow \phi)$ becomes provable in KF with this definition. The schema had been mentioned by Kripke [28] already.

Reinhardt [37, 38] takes the determinateness predicate to express *meaningful applicability* or *significance*; and truth is only meaningfully applicable to determinate sentences.[13] In particular, the liar sentence is not in the range of significance of the truth predicate. Reinhardt concludes that the system KF, which is formulated in classical logic, is not sound in the sense that it proves only true theorems, where truth is understood in the sense of the truth predicate T of KF. According to Reinhardt, one might hope that the sentences ϕ with $\mathsf{KF} \vdash \mathrm{T}\ulcorner\phi\urcorner$ are true and significant. The set of these sentences is not closed under classical logic. The theorems of KF itself need not be significant or trustworthy; only those whose truth can be proved are. Generally, it is very hard to avoid the provability of sentences that are not 'significant' or not 'healthy', as Reinhardt [37] and Bacon [1] have argued.

In contrast to Reinhardt, we do not think that the range of significance of the truth predicate is restricted in any way; the determinateness predicate cannot be seen as indicating the range of significance of T. For instance, truth should commute with connectives for *all* sentences, not just with those in some range of significance. However, some sentences, including liar sentences, are sensitive to the addition of another layer of truth. Stacking an additional layer of truth onto the liar sentence will change its semantic status; but that does not mean that the truth predicate cannot meaningfully applied to it. We would only run into the Reinhardt–Bacon problem by postulating that only insensitive sentences ought to be provable. We reject this kind of soundness condition. We endorse sensitive sentences such as the compositional axioms or classical tautologies as theorems. However, we may not be able to ascend semantically from such sentences and add a further level of truth. It is not possible to ascend semantically from the compositional axiom; they are already at the highest level of generality and further semantic ascend and generalization is not possible.

10.3. Comparison with Feferman's DT. Feferman [10, p. 205] originally introduced KF as an instrument to explain 'what notions and principles one ought to accept if one accepts the basic notions and principles of the theory', the foundational question which had long been one of the central themes of Feferman's work.[14]

As a theory of truth *per se*, Feferman later proposed another system DT. Again, each predicate is assumed to have a *domain of significance* and to be meaningfully applicable only to objects in that domain. In the case of the truth predicate T, its domain $\mathfrak{D}$ of significance is taken to consist of the sentences that are *meaningful*

[12]The system is not called KF in [8]. Moreover, different systems have been called KF. See Halbach [21] for more on the history of KF and its variants.

[13]That predicates have a domain of significance has been part of the philosophy of type theory since [39]. Feferman [10] explicitly refers to Russell.

[14]However, Feferman [10, p. 205] expressed 'I always thought that the KF axioms were a bit artificial for that purpose' and abandoned KF as such an instrument in the end. In place of KF, he proposed a new notion of *unfolding* of a schematic system in place of KF for the purpose in question.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

and determinate, and all the principles of truth are only applicable to the sentences in the domain $\mathfrak{D}$. Accordingly, the system DT comprises two groups of axioms: the axioms characterizing the domain $\mathfrak{D}$ and the axioms expressing the T-schema and the compositional axioms for all sentences in $\mathfrak{D}$.

Feferman's axioms for determinateness differ from ours in particular with respect to the determinateness of sentences formed with binary connectives. In his system a disjunction is determinate if, and only if both disjuncts are; and a conjunction is determinate if, and only if both conjuncts are. We discussed our choice of axiom D5 on page 6 and justified its endorsement with the generalizing function of truth. Many harmless generalizations would become indeterminate (and consequently the system proof-theoretically weak) if generalization were expressed with conjunction and disjunction and Feferman's concept of determinateness. To overcome the problem, he introduced a special conditional $\rightarrow$ as a primitive logical connective so that the truth and determinateness of a conditional sentence $\phi \rightarrow \psi$ is characterized independently and differently from its usual definition $\neg(\phi \wedge \neg\psi)$ in terms of negation and conjunction (or $\neg\phi \vee \psi$ in terms of negation and disjunction) by the following axioms:

$$\forall x\, \forall y \left(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (\mathfrak{D}(x \dot{\rightarrow} y) \leftrightarrow (\mathfrak{D}x \wedge (\mathrm{T}x \rightarrow \mathfrak{D}y)))\right),$$
$$\forall x\, \forall y \left(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \wedge \mathfrak{D}(x \dot{\rightarrow} y) \rightarrow (\mathrm{T}(x \dot{\rightarrow} y) \leftrightarrow (\mathrm{T}x \rightarrow \mathrm{T}y))\right).$$

Feferman's approach diverges from ours in two crucial respects: First, he took $\mathfrak{D}$ as the range of all the principles of truth comprehensively and restricted every principle of truth to the class of determinate sentences $\mathfrak{D}$, whereas we take the class of determinate sentences only as the range of the disquotation schema and postulate the compositional axioms unrestrictedly for every sentence. Secondly, he took $\mathfrak{D}$ as definable in terms of truth by stipulating $Dx \leftrightarrow \mathrm{T}x \vee \mathrm{F}x$; but this definition does not yield the desired properties of D in our theory and we consequently introduce determinateness as a primitive notion.

Fujimoto [14] observed that DT is identical with the system FKF + Cons of non-classical partial truth whose evaluation rule is given by the Aczel–Feferman evaluation schema. The Aczel–Feferman evaluation schema is the same as the weak Kleene evaluation schema, with the exception of the evaluation rule of the conditional. FKF is a variant of KF with the Aczel–Feferman evaluation schema: FKF is to the Aczel–Feferman evaluation schema is what KF is to the strong Kleene evaluation schema.

10.4. Comparison with the Leitgeb–Schindler system. Schindler [40] defined a group of systems that resemble ours. They differ from CD in restricting the compositional principles T4 –T6 to *grounded* sentences. The predicate symbol G for groundedness plays a role comparable to that of D in CD. The full system is called CG. All the unrestricted compositional axioms T4–T6 are provable except for the right-to-left direction of the negation axiom T4. That is, $\forall x \left(\mathrm{Sent}(x) \rightarrow (\neg\mathrm{T}x \rightarrow \mathrm{T}(\dot{\neg}x))\right)$ is not provable in CG. In this sense Schindler's system fails to be thoroughly classical.

One of the axioms of CG is a 'definitional' axiom of groundedness [40, p. 77]:

$$\forall x\, (Gx \leftrightarrow (\mathrm{T}x \vee \mathrm{T}\dot{\neg}x)).$$

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

This axiom of CG is not a theorem (with D for G) of our theory CD, because the full negation axiom T4 and thus $\forall x\,(\mathrm{Sent}(x) \to (\mathrm{T}x \vee \mathrm{T}\dot{\neg}x))$ are theorems of CD and therefore Schindler's definitional axiom would imply $\forall x\,(\mathrm{Sent}(x) \to Gx)$. Consequently, Schindler's system does not really require a primitive predicate for groundedness, while the definition of determinateness as truth or falsity fails in our system.

On Schindler's approach, T applies provably only to sentences without G, because it lacks an axiom analogous to our T2. This is in line with the intended interpretation of G as Leitgeb's [30] notion of groundedness. Also G cannot be iterated, and axiom D3 is absent from Schindler's list of axioms. However, contrary to Schindler's official formulation of the system, we could view Gx as a metalinguistic abbreviation for $\mathrm{T}x \vee \mathrm{T}\dot{\neg}x$. Then T provably applies to sentences containing G.

10.5. Comparison with Halbach's system PUTB. The truth-theoretic axioms of the system PUTB are given by all instances $\mathrm{T}\ulcorner\phi\urcorner \leftrightarrow \phi$ of the T-schema where the truth predicate occurs only positively in ϕ and parameters are allowed in ϕ.

Positiveness has the advantage over determinateness that it is a very simple syntactic decidable property that is definable in the language of arithmetic. If in the schema

$$\mathsf{DDS} \qquad \forall t_1 \ldots \forall t_n \left(\mathrm{D}\ulcorner\phi(\dot{t}_1, \ldots, \dot{t}_n)\urcorner \to \left(\mathrm{T}\ulcorner\phi(\dot{t}_1, \ldots, \dot{t}_n)\urcorner \leftrightarrow \phi(t_1{}^\circ, \ldots, t_n{}^\circ)\right)\right),$$

the symbol D expresses that T occurs only positively, no primitive predicate D is needed, as positiveness can be expressed by an arithmetical formula. Consequently, because D is arithmetical, our axiom $\mathsf{T2}^+$, that is, $\forall s(\mathrm{T}\dot{\mathrm{D}}s \leftrightarrow \mathrm{D}s^\circ)$ becomes provable.

The theory PUTB does not prove the compositional axiom T4 –T6. To obtain a system closer to CD, one could add the compositional axioms to PUTB or derive them from reflection principles, as Horsten and Leigh [27] suggest.

However, T-positiveness and determinateness differ significantly: In particular, truth-teller sentences are provably T-positive, while they are not provably determinate in our sense; in fact we can refute their determinateness in CD_μ as was shown in proposition 9.3. The proof-theoretic strength of PUTB relies crucially on indeterminate instances that allow one to mimick positive inductive definitions. Overall the restriction to determinate instead of positive sentences is better motivated.

10.6. Comparison with Picollo's system WFUTB. Picollo [34, 35] defines a notion of referential well-foundedness, which is related to determinateness in our sense. Here we do not go into the details of its definition and only highlight some differences to our approach. First, the notion of well-foundedness is defined in such a way that it becomes arithmetically definable. Thus only a truth predicate needs to be added and no separate predicate for referential well-foundedness as our primitive predicate D. Roughly, the main axiom schema of her system WFUTB then states the T-sentences for sentences that are referentially well-founded in her sense (or PA- provably equivalent to such a sentence).[15] WFUTB does not feature

[15] Strictly, speaking there are two conditions on the permissible instances of the T-schema: The uniform T-sentences are postulated for instances that are provably r-stable and well-founded in PA. By *referential*

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

any compositional axioms, and it can be shown that they are not provable in it, although the system was shown by Picollo to be proof-theoretically at least as strong as ramified analysis $\mathsf{RA}_{<\Gamma_0}$ up to Γ_0 and thus much stronger than PUTB and compositional theories such as KF.

10.7. Comparison with Field's systems Int. The most fundamental decision for the truth theorist is whether to sacrifice classical logic for transparency or transparency for classical logic. By transparency we mean here some equivalence of ϕ with $\mathrm{T}^{\ulcorner}\phi^{\urcorner}$ for all ϕ. Field [11] sacrifices classical logic; we sacrifice transparency. Field saves truth from paradox; we save logic from paradox. We regain transparency for determinate sentences; Field regains classical logic for what he calls *strongly classical* sentences.

Field employs a new primitive predicate *Scl* for strongly classical truth. Although Field's and our approach are pulling in exactly opposite directions, Field's axioms for *Scl* and ours for D have a striking resemblance. As Field in footnote 6 mentions, he and we arrived at our axiomatizations independently. Of course, a serious comparison of Int and CD leads back to the most fundamental decision that truth theorists face, and we do not enter the discussion here.

§11. Further perspectives. In the part II of this paper, we will give proof-theoretic analysis of variants of CD. In particular, among many others, we will give a proof of Theorem 9.4. We conclude the paper by listing two open problems.

(I) We may consider CD and its variants with the schema of induction restricted to the arithmetical sentences. In many cases, the restriction of induction to the arithmetical sentences results in a proof-theoretically conservative theory (over PA). We conjecture that the same holds for CD. Conservativeness proofs in the analogous case of KF can be given in a model-theoretic way by showing how to extend a given model of PA to a model of KF with arithmetical induction only. This is not possible in the case of CD because Lachlan's theorem [29] applies and nonstandard models that can be expanded to CD with restricted induction have to be recursively saturated.

(II) It may be of interest to replace the axiom D3 of CD with an alternative axiom $\forall s\, \mathrm{D}\dot{\mathrm{D}}s$. Together with DDS this would yield our additional axiom $\mathsf{T2}^{+}$. Thus, adding $\forall s\, \mathrm{D}\dot{\mathrm{D}}s$ may be more natural than adding $\mathsf{T2}^{+}$. We do not know how strong CD becomes if $\forall s\, \mathrm{D}\dot{\mathrm{D}}s$ is added.

Acknowledgments. We are grateful to Dora Achourioti, Anton Broberg, Hannes Leitgeb, Beau Mount, Carlo Nicolai, Lavinia Picollo, Lorenzo Rossi, Thomas Schindler, Johannes Stern, Philip Welch, and the referee for comments and suggestions.

Funding. V.H. would like to thank the Leverhulme Trust for supporting his work with a Research Fellowship.

well-foundedness we mean the conjunction of the two conditions. Unlike most notions of groundedness and determinateness, referential well-foundedness is sensitive to the base theory.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

REFERENCES

[1] A. Bacon, *Can the classical logician avoid the revenge paradoxes?* ***Philosophical Review***, vol. 124 (2015), pp. 299–352.

[2] E. Barrio and L. Picollo, *Notes on ω-inconsistent theories of truth in second-order languages*. ***Review of Symbolic Logic***, vol. 6 (2013), pp. 733–741.

[3] J. C. Beall and B. Armour-Garb, ***Deflationism and Paradox***, Clarendon Press, Oxford, 2005.

[4] J. P. Burgess, *Friedman and the axiomatization of Kripke's theory of truth*, ***Foundational Adventures: Essays in Honor of Harvey M. Friedman***, 2009, paper delivered at the Ohio State University conference in honor of the 60th birthday of Harvey Friedman, http://foundationaladventures.com/.

[5] A. Cantini, *Notes on formal theories of truth*. ***Zeitschrift für mathematische Logik und Grundlagen der Mathematik***, vol. 35 (1989), pp. 97–130.

[6] ———, *A theory of formal truth arithmetically equivalent to* ID_1, this Journal, vol. 55 (1990), pp. 244–259.

[7] S. Feferman, *Hilbert's program relativized: Proof-theoretical and foundational reductions*, this Journal, vol. 53 (1988), pp. 364–384.

[8] ———, *Reflecting on incompleteness,* this Journal, vol. 56 (1991), pp. 1–49.

[9] S. Feferman, *What rests on what? The proof-theoretic analysis of mathematics*, ***Proceedings of the 15th International Wittgenstein Symposium***, Hölder-Pichler-Tempsky, Wien, 1992, pp. 47–171.

[10] S. Feferman, *Axioms for determinateness and truth*. ***Review of Symbolic Logic***, vol. 1 (2008), pp. 204–217.

[11] H. Field, *The power of naive truth*. ***Review of Symbolic Logic***, (2020), pp. 1–34.

[12] M. Fischer, L. Horsten, and C. Nicolai, *Hypatia's silence*. ***Noûs***, vol. 55 (2021), pp. 62–85.

[13] H. Friedman and M. Sheard, *An axiomatic approach to self-referential truth*. ***Annals of Pure and Applied Logic***, vol. 33 (1987), pp. 1–21.

[14] K. Fujimoto, *Relative truth definability*. ***Bulletin of Symbolic Logic***, vol. 16 (2010), pp. 305–344.

[15] ———, *Notes on some second-order systems of iterated inductive definitions and* Π^1_1*-comprehensions and relevant sybsystems of set theory*. ***Annals of Pure and Applied Logic***, vol. 166 (2015), pp. 409–463.

[16] ———, *Deflationism beyond arithmetic*. ***Synthese***, vol. 196 (2019), pp. 1045–1069.

[17] ———, *The function of truth and the conservativeness argument*. ***Mind***, vol. 131 (2022), no. 521, pp. 129–157.

[18] V. Halbach, *A system of complete and consistent truth*. ***Notre Dame Journal of Formal Logic***, vol. 35 (1994), pp. 311–327.

[19] ———, *Disquotational truth and analyticity,* this Journal, vol. 66 (2001), pp. 1959–1973.

[20] ———, *How not to state the T-sentences*. ***Analysis***, vol. 66 (2006), pp. 276–280.

[21] ———, ***Axiomatic Theories of Truth***, revised ed., Cambridge University Press, Cambridge, 2014 (first edition 2011).

[22] ———, *Formal notes on the substitutional analysis of logical consequence*. ***Notre Dame Journal of Formal Logic***, vol. 61 (2020), pp. 317–339.

[23] ———, *The substitutional analysis of logical consequence*. ***Noûs***, vol. 54 (2020), pp. 431–450.

[24] V. Halbach and G. Leigh, ***The Road to Paradox: A Guide to Syntax, Truth, and Modality***, Cambridge University Press, 2024, to appear.

[25] J. van Heijenoort (editor), ***From Frege to Gödel***, A Source Book in Mathematical Logic, 1879–1931, Harvard University Press, Cambridge, 1967.

[26] W. Hodges, *Truth in a structure*. ***Proceedings of the Aristotelean Society***, vol. 86 (1986), pp. 135–151.

[27] L. Horsten and G. E. Leigh, *Truth is simple*. ***Mind***, vol. 126 (2017), pp. 195–232.

[28] S. Kripke, *Outline of a theory of truth*. ***Journal of Philosophy***, vol. 72 (1975), pp. 690–716.

[29] A. Lachlan, *Full satisfaction classes and recursive saturation*. ***Canadian Mathematical Bulletin***, vol. 24 (1981), pp. 295–297.

[30] H. Leitgeb, *What truth depends on*. ***Journal of Philosophical Logic***, vol. 34 (2005), pp. 155–192.

[31] V. McGee, *How truthlike can a predicate be? A negative result*. ***Journal of Philosophical Logic***, vol. 14 (1985), pp. 399–410.

[32] ———, *Maximal consistent sets of instances of Tarski's schema (T)*. ***Journal of Philosophical Logic***, vol. 21 (1992), pp. 235–241.

[33] L. Picollo, *Reference in arithmetic*. ***Review of Symbolic Logic***, vol. 11 (2018), pp. 573–603.

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press

[34] ———, *Reference and truth*. ***Journal of Philosophical Logic***, vol. 49 (2020), pp. 439–474.

[35] ———, *Minimalism, reference, and paradoxes*, ***The Logica Yearbook 2015*** (P. Arazim and M. Dancak, editors), College Publications, London, 2016, pp. 163–178.

[36] W. Pohlers, ***Subsystems of set theory and second order number theory***, ***Handbook of Proof Theory*** (S. Buss, editor), Studies in Logic and the Foundations of Mathematics, 137, Elsevier, Amsterdam, 1998, pp. 209–335.

[37] W. Reinhardt, *Some remarks on extending and interpreting theories with a partial predicate for truth*. ***Journal of Philosophical Logic***, vol. 15 (1986), pp. 219–251.

[38] ———, ***Remarks on significance and meaningful applicability***, ***Mathematical Logic and Formal Systems: A Collection of Papers in Honor of Professor Newton C.A. Da Costa*** (L. P. de Alcantara, editor), Lecture Notes in Pure and Applied Mathematics, 94, Marcel Dekker, 1985, pp. 227–242.

[39] B. Russell, *Mathematical logic as based on the theory of types*. ***American Journal of Mathematics***, vol. 30 (1908), pp. 222–262, Reprinted in [25, 150–182].

[40] T. Schindler, *Axioms for grounded truth*. ***Review of Symbolic Logic***, vol. 7 (2014), pp. 73–83.

[41] J. Stern, *Supervaluation-style truth without supervaluations*. ***Journal of Philosophical Logic***, vol. 47 (2018), pp. 817–850.

[42] A. Tarski, *Der Wahrheitsbegriff in den formalisierten Sprachen*. ***Studia Philosophica Commentarii Societatis Philosophicae Polonorum***, vol. 1 (1935), pp. 261–405.

[43] A. Tarski and R. Vaught, *Arithmetical extensions of relational systems*. ***Compositio Mathematica***, vol. 13 (1956), pp. 81–102.

NEW COLLEGE
UNIVERSITY OF OXFORD
OXFORD OX1 3BN
UK
E-mail: volker.halbach@new.ox.ac.uk

SCHOOL OF MATHEMATICS
UNIVERSITY OF BRISTOL
WOODLAND ROAD
BRISTOL BS8 1UG
UK
E-mail: kentaro.fujimoto@bristol.ac.uk

https://doi.org/10.1017/jsl.2023.49 Published online by Cambridge University Press