Erschienen in: Journal of Ecology. Wiley. 2025, 113(10), S. 3019-3036
https://doi.org/10.1111/1365-2745.70146

Truth and Finite Conjunction

Guanglong Luo
Nankai University, China
glluo@nankai.edu.cn

Leon Horsten
University of Konstanz, Germany
leon.horsten@uni-konstanz.de

Sam Roberts
University of Konstanz, Germany
sam.roberts@uni-konstanz.de

This note is a critical response to Kentaro Fujimoto's new conservativeness argument about truth, which centres on the notion of finite conjunction. We argue that Fujimoto's arguments turn on a specific way of formalizing the notions of finite collection and finite conjunction in first-order logic. In particular, by instead formalizing these concepts in a natural way in set theory or in second-order logic, Fujimoto's new conservativeness argument can be resisted.

1. Deflationism and conservativeness

According to the central claim of *deflationism about truth*, the concept of truth is a 'light', 'insubstantial' notion. This raises the question what kind of notion truth is instead. Many deflationists take truth somehow to be a *quasi-logical* notion. Just as it makes no sense to ask what the essence of 'and' is, so it makes no sense to ask what the essence of truth is. Rather, like 'and', truth is a notion that helps us reason well.

The claim that truth is an insubstantial notion is rather vague; it can be made clearer in several ways. One way in which the claim that truth is an insubstantial notion can be made more precise is in terms of *proof-theoretic conservativeness*. On this precisification, a notion of truth that is governed by a set of axioms Tr is deflationist if for any theory S that does not explicitly involve the notion of truth, the extension of S with the axioms of Tr—call this theory $Tr[S]$—is proof-theoretically conservative over S. Here $Tr[S]$ is said to be proof-theoretically conservative over S if and only if $Tr[S]$ proves no *new* theorems in the language of S: if a theorem in the language of S is provable in $Tr[S]$, then it is already provable in S alone (that is, without making use of the concept of truth).

Conservativeness deflationism has been implicitly or explicitly endorsed by several authors. Horsten and Leigh (2017), for instance, defend a conservative type-free disquotational truth theory as our fundamental theory of truth, and thus indirectly endorse conservativeness deflationism. Waxman explicitly endorses conservativeness deflationism in the following passage:

> Is there any reasonable scope for denying that a deflationist theory of truth must be conservative? … The transition [from the claim that truth is insubstantial to the claim that truth is conservative] has considerable intuitive force, for it seems extremely uncomfortable to maintain that truth is an insubstantial or non-robust property if the addition of truth principles leads one to rule out what were previously considered to be live possibilities concerning a (truth-free) subject matter. Perhaps the best way of understanding the transition is as a proposed explication: the informal notion of metaphysical insubstantiality is to be (possibly partially) explicated in terms of the formal criterion of conservativeness. It is striking, and a mark in favour of the plausibility of this understanding, that the conservativeness requirement has attracted considerable support among deflationists themselves. (Waxman 2017, pp. 445–6)

Conservativeness deflationism can also be made more precise in terms of other concepts, such as semantic conservativeness. However, the proof-theoretic explication is often presupposed in the debate about the viability of deflationism about truth. It plays a central role in the so-called *conservativeness argument* against deflationism of truth, which goes as follows. A powerful and seemingly unobjectionable theory of truth is the compositional truth theory $\mathsf{CT}|\mathsf{CT}[S]$, according to which the truth predicate commutes with the first-order logical connectives. It can be shown that under fairly general conditions, $\mathsf{CT}[S]$ is *not* proof-theoretically conservative over a background theory S when the schematic commitments of S (such as its logical commitments, and its commitment to the principle of mathematical induction) are taken in an open-ended way rather than restricted to the language in which S is formulated. *Therefore*—or so the argument goes—truth is a substantial notion, and deflationism about truth is false.

Even given the proof-theoretic explication of deflationism about truth, the conservativeness argument is controversial. Perhaps the most influential objection against it was raised by Field (1999).[1] He argues that extending mathematical induction so as to include predicates containing the truth predicate amounts to a mathematical strengthening of

---

[1] For a recent critical discussion conservativeness deflationism, see Murzi and Rossi (2020).

the background theory. If this is right, then the conservativeness argument is blocked. Let us denote the result of adding the compositional axioms for truth *without extending the schematic mathematical axioms* of a background theory S as $\mathsf{CT}^{-}[S]$ (while continuing to reserve the designation $\mathsf{CT}[S]$ for the result of extending S with the compositional truth axioms *and* extending all its schemes to allow instances of the truth predicate). Then $\mathsf{CT}^{-}[S]$ is proof-theoretically conservative over S, because inductive arguments concerning formulae including the truth predicate cannot then be carried out in $\mathsf{CT}^{-}[S]$.

Fujimoto (2022) formulates and discusses a *new conservativeness argument*. Like the 'old' conservativeness argument, the new argument also intends to establish the proof-theoretic non-conservativeness of truth, and thereby refute deflationism about truth. But it is intended to be convincing even to those who accept Field's critique of the old conservativeness argument.

The basic idea of the new conservativeness argument is that $\mathsf{CT}^{-}[S]$ fails to fully capture the compositionality of the truth predicate. In particular, $\mathsf{CT}^{-}[S]$ does not prove that any *arbitrary finite* conjunction of (truth-free) statements is true if and only if every one of these statements is true. This property of truth is called the principle of *conjunctive correctness*. Fujimoto produces informal arguments the validity of which turns on instances of the principle of conservative correctness. Thus, Fujimoto argues, the truth theory $\mathsf{CT}^{-}[S]$ should be extended by the principle of conjunctive correctness. But it has been shown in recent years that if $\mathsf{CT}^{-}[S]$ is extended by a particular formalization of conjunctive correctness, then the resulting theory $\mathsf{CT}^{cc}[S]$ is under fairly general conditions proof-theoretically non-conservative over S *even if no predicates containing the truth predicate are allowed in arguments by mathematical induction*. Thus deflationism of truth fails even if Field's critique of the 'old' conservativeness argument is correct.

We shall see how Fujimoto's argument turns out to be quite sensitive to the way in which conservative correctness is formally expressed. Fujimoto formalizes conjunctive correctness in a first-order setting, in terms of an arithmetical coding of finite sets and an elementhood relation on finite sets. If instead the notion of finite set is formalized in a natural way in second-order logic or set theory, then the resulting truth theory is proof-theoretically conservative.

We will argue that the straightforward formalization of finiteness in second-order logic has an intuitive appeal that Fujimoto's formalization involving coded expression of finite sets in first-order arithmetic lacks.

One can argue about the advantages and disadvantages of formalizing conjunctive correctness in first-order arithmetic, set theory, or second-order arithmetic. But it should not be a commitment *of truth theory* that the concepts of finite set and of finite conjunction are expressed in one particular setting. Therefore Fujimoto's new conservativeness argument should not sway even the philosophers who accept the proof-theoretic explication of deflationism about truth.

2. Blind deduction

Following Fujimoto, we work in a typed setting, so that the discussion is unaffected by complications induced by self-applicable truth. Moreover, in what follows, as is often done, for concreteness we mostly take the background theory S to be PA. But the arguments that we give are intended to apply more generally to compositional truth applied to arbitrary background theories, as long as they are mathematically sufficiently strong. Indeed, in §4.1 we briefly discuss the situation where first-order set theory, rather than first-order arithmetic, is taken as the background theory. Moreover, we will consider the prospects of appealing to second-order resources in the formalization of arguments that involve the concept of arbitrary finite conjunction.

Basic truth principles are needed to explain why certain apparently valid informal arguments using the concept of truth are indeed valid. Fujimoto argues that for truth theories to fulfil this task adequately, they must at least contain the compositional truth axioms (Fujimoto 2022, §3). In other words, he argues that the truth theory CT^{-} should be taken for granted. We do not challenge this assumption in this note.

Truth is used in what Fujimoto calls *blind deductions*, which are 'deductive arguments about the truth of some sentences by analysing and manipulating their logico-linguistic structure without explicitly specifying what these sentences are' (Fujimoto 2022, p. 137).

Examples of blind deductions play a key role in Fujimoto's new conservativeness argument. In particular, Fujimoto makes use in his new conservativeness argument of the following three pieces of informal reasoning:

ARGUMENT 1 (Fujimoto 2022, p. 147)

(P1) All the axioms of PA are true.

(P2) Amy wrote down some (finitely many) axioms of PA in her notebook.

(P3) If what Amy wrote down in her notebook is all true, then Cathy's conjecture is true.
(P4) Cathy made exactly one conjecture.
(C1) Cathy's conjecture is true.

Argument 2 (Fujimoto 2022, p. 149)
(P1), (P2) (The first two premisses of Argument 1.)
(P5) Beth denied one of the sentences that Amy wrote in her notebook.
(P6) If Cathy's conjecture is true, then Beth's claim is true.
(P7) Beth made exactly one claim and Cathy made exactly one conjecture.
(C2) If what Amy wrote in her notebook is all true, then Cathy's conjecture is not true.

Argument 3 (Fujimoto 2022, p. 151)
(P1), (P2), (P7) (Premisses 1, 2, and 7.)
(P8) Beth claimed *the conjunction of* what Amy wrote in her notebook implies Cathy's conjecture.
(C3) If Beth's claim is true, then Cathy's conjecture is true.

Observe that Argument 3 can be seen as a 'variant' of Argument 1.

Let us violate Russellian strictures about the formalization of descriptions slightly by formalizing 'Cathy's conjecture' as a (first-order) individual constant *a* and 'Beth's claim' as a (first-order) individual constant *b*. This simplifies matters—since we can then ignore premisses P4 and P7—without affecting the structure of the argument (as the reader can readily verify).

Fujimoto claims that all three arguments are intuitively valid (2022, pp. 147, 149, 150), and so do we. If truth is a quasi-logical notion, then an adequate axiomatic theory must bear this out, by being such that from correct formalizations of the premisses, the conclusions can be *derived* using truth axioms.

3. From blind deduction to conjunctive correctness

The concept of finiteness seems to play a role in all three arguments. In fact, we will see that it is not clear that the concept of finiteness plays an essential role in the first two arguments; but is a crucial ingredient in Argument 3.

The following is a first-order way of making sense of the concept of finiteness in play. There is an arithmetical expression $\underset{\cdot}{\in}$ belonging to the language of first-order Peano Arithmetic ($\mathcal{L}_{\mathsf{PA}}$) such that for every finite set of (codes of) sentences X, there is an arithmetical code c of X such that, for all natural numbers n,

$$n \in X \Leftrightarrow n \underset{\cdot}{\in} c.$$

Moreover, PA proves (a coded version of) comprehension for finite predicates (Fujimoto 2022, p. 145):

> *Definition* 1 (FC). $\Phi(x)$ has a finite extension, that is, $\exists n \forall x(\Phi(x) \rightarrow x \leq n)$ if and only if $\exists c \forall x(x \underset{\cdot}{\in} c \leftrightarrow \Phi(x))$.

In addition, there is an arithmetically definable function $\underset{\cdot}{\wedge}$ that transforms a code c of a finite set of sentences into the conjunction of these sentences (with a given bracketing convention). This then gives us a notion of *blind conjunction* ($\underset{\cdot}{\wedge} x$).

Now for any given a finite collection X of sentences, Fujimoto claims that the truth of the blind conjunction of the Xs should be formalized as

$$T(\underset{\cdot}{\wedge} c),$$

with c the code of X, and T a primitive truth predicate (Fujimoto 2022, p. 146).

Then the following principle, which is known as the axiom of *Conjunctive Correctness*, can be formulated:

Axiom 1 (CC).

$$\forall c : (\forall x(x \underset{\cdot}{\in} c \rightarrow x \in \mathcal{L}_{\mathsf{PA}})) \rightarrow ((\forall x\,(x \underset{\cdot}{\in} c \rightarrow T(x))) \leftrightarrow T(\underset{\cdot}{\wedge} c)).$$

The version of CC with the consequent restricted to a left-to-right implication is known as $\mathrm{CC}_{\mathrm{Intro}}$. The version of CC with the consequent restricted to a right-to-left implication is known as $\mathrm{CC}_{\mathrm{Elim}}$.

Define $\mathsf{CT}^{\mathrm{cc}}[\mathsf{PA}]$ as the theory resulting from adding the axiom CC to $\mathsf{CT}^{-}[\mathsf{PA}]$. Enayat and Pakhomov (2019) have proved the following surprising theorem:

> *Theorem* 2. $\mathsf{CT}^{\mathrm{cc}}[\mathsf{PA}] \vdash \mathsf{CT}_0[\mathsf{PA}]$,

where $\mathsf{CT}_0[\mathsf{PA}]$ is like $\mathsf{CT}^{-}[\mathsf{PA}]$, except that the induction axioms for *quantifier-free* atomic formulae *that may contain occurrences of the truth predicate* are also included. Now, Wcisło and Łełyk (2017) have shown that $\mathsf{CT}_0[\mathsf{PA}]$ is arithmetically non-conservative over PA. So this means that $\mathsf{CT}^{\mathrm{cc}}[\mathsf{PA}]$ is also arithmetically non-conservative over PA.

With all this in place, Fujimoto argues that the conclusions of Arguments 1 and 2 can be derived from their premisses *only if* CC holds.

More specifically, in the context of $\mathsf{CT}^-[\mathsf{PA}]$ Argument 1 is a derivable argument scheme only if $\mathrm{CC}_{\mathrm{Intro}}$ holds, and Argument 2 is a derivable argument scheme only if $\mathrm{CC}_{\mathrm{Elim}}$ holds (Fujimoto 2022, §4). We do not rehearse his argument here, but merely stress that his argument heavily depends on formalizing these arguments using the machinery of coding finite sequences in first-order arithmetic in the way described above. Since in the context of $\mathsf{CT}^-[\mathsf{PA}]$, CC is arithmetically non-conservative over the background arithmetical theory PA, Fujimoto concludes that truth is non-conservative.

4. Truth, finiteness, and second-order logic

We now turn to the evaluation of Fujimoto's *new* conservativeness argument. All three arguments are intended to indicate that conjunctive correctness should be added to $\mathsf{CT}^-[\mathsf{PA}]$ as a fundamental truth axiom, and all three arguments are somehow connected with the notions of finiteness. We accept that a form of conjunctive correctness needs to be provable from our basic principles governing the notion of truth. In the following, we critically evaluate the role and formal treatment of finiteness in Fujimoto's three arguments.

4.1. In set theory

The notion of finiteness is of course straightforwardly expressible in the language of set theory ($\mathcal{L}_{\mathsf{ZFC}}$). So suppose we take first-order ZFC as our background theory, and—as in the arithmetical case—add compositional truth axioms to it, but be careful *not* to allow the truth predicate to occur in the separation and replacement schemes. Call the resulting theory $\mathsf{CT}^-[\mathsf{ZFC}]$. Then we can define a natural (coding-free) notion of corrective correctness in the following manner:

> *Definition* 3 (CC^{set}). $\forall x : [|x| < \omega \wedge \forall y \in x : y \in \mathcal{L}_{\mathsf{ZFC}}] \rightarrow [T(conj(x)) \leftrightarrow \forall y \in x : Ty]$.

It is clear that in the theory $\mathsf{CT}^-[\mathsf{ZFC}] + \mathrm{CC}^{set}$, the obvious formalizations of the three arguments can be proved.

However, it follows from an argument by Fujimoto (2012, Theorem 20) that:

> *Theorem* 4. $\mathsf{CT}^-[\mathsf{ZFC}] + \mathrm{CC}^{set}$ is conservative over ZFC for the language of set theory.

Thus, as Fujimoto himself notes (2022, p. 155), his new conservativeness argument does not go through in this setting.

What is wrong with formalizing the three arguments in the setting of set theory? One possible worry might be that the ontological commitments of set theory far outstrip the ontological commitments of the three arguments. It might seem, in other words, that the price for ideological conservativeness is ontological non-conservativeness. For this reason, we shall now attempt to show that even in a setting that is ontologically conservative over first-order arithmetic, the three arguments do not force non-conservativeness of truth upon us.

4.2. *The first two arguments*

The qualification 'finitely many' is in brackets in Argument 1 and implicitly assumed in Argument 2—witness Fujimoto's formalization of Argument 2 in Fujimoto (2022, p. 149)—so it is somewhat ambiguous whether it belongs to the argument. If we ignore the qualification 'finitely many' in our formalization (and therefore do not need the first-order machinery of coded finite sets at all), then the validity of Arguments 1 and 2 can be witnessed in the background theory alone or in $\mathsf{CT}^{-}[\mathsf{PA}]$. So, in that case, non-conservativeness does not follow from these arguments. Let us formalize Argument 1 in this way, where the predicate N formalizes 'is written down by Amy' (and, as said before, a is an individual constant referring to Cathy's conjecture):

(1) $\forall x : Ax\mathsf{PA}(x) \rightarrow Tx$
(2) $\forall x : N(x) \rightarrow Ax\mathsf{PA}(x)$
(3) $(\forall x : N(x) \rightarrow T(x)) \rightarrow T(a)$
(4) $T(a)$

It is immediate that the last sentence is derivable from the previous ones. Indeed, truth laws play no role in this derivation. Thus the parenthetical finiteness assumption appears to be a red herring. Moreover, we do not even need the truth laws of $\mathsf{CT}^{-}[\mathsf{PA}]$ in this derivation. This depends on formalizing '*A* implies *B*' as $T(A) \rightarrow T(B)$.[2] Alternatively, one could formalize '*A* implies *B*' as $T(A \rightarrow B)$. Then some of the compositional truth axioms of $\mathsf{CT}^{-}[\mathsf{PA}]$ play a role in the derivation. In either case, on this reading of Argument 1, we do not obtain proof-theoretic non-conservativeness.[3]

---

[2] Actually, it should probably rather be formalized as '*Necessarily*, if *A* is true, then *B* is true'. But, like Fujimoto, we ignore the modal aspect of implication in this note.

[3] A completely parallel analysis can be given of Argument 2. We leave this analysis to the reader.

As is well known, the notion of finiteness can be explicitly *defined* in a simple and natural way in second-order logic. Moreover, since second-order logic can be interpreted in an *ontologically* non-inflationary way as plural logic,[4] we take it to be in principle philosophically unobjectionable to make use of second-order logic for the purposes of formalization of natural language arguments.

If we do build the parenthetical finiteness claims into our formalization, but formalize finiteness in a second-order setting, then the conclusions of Fujimoto's first two arguments again follow from their premisses very directly. Let $FIN(X)$ be a standard second-order definition of what it means for X to be finite. Then Argument 1, for instance, can be formalized in the language of second-order arithmetic as follows:[5]

(1) $\forall x : Ax\mathsf{PA}(x) \rightarrow Tx$
(2) $\exists X[FIN(X) \wedge \forall y : N(y) \leftrightarrow (y \in X \wedge Ax\mathsf{PA}(y))]$
(3) $(\forall x : N(x) \rightarrow T(x)) \rightarrow T(a)$
(4) $T(a)$

For the last sentence to be derivable from the premisses, it suffices to derive $\forall x : N(x) \rightarrow T(x)$ from the first two premisses. But this can easily be done using just the normal existential generalization and instantiation rules for second-order logic,[6] and without using truth laws. Again, the bit about finiteness in the formalization plays no active role in the derivation: it is a red herring. Note that in particular, therefore, no use is made of any kind of second-order comprehension (or mathematical induction) in this derivation. This means that the whole derivation can easily take place in a second-order theory such as ACA_0, which is first-order conservative over PA (and even this is overkill).[7]

4.3. The third argument

Fujimoto's Argument 3 is subtle. According to P8, Beth does not make a claim concerning any *specific* conjunction of statements: Beth makes a *de dicto* rather than a *de re* claim. This is the reason why the concept of *arbitrary* finite conjunction is needed to formalize the argument.[8]

---

[4] See, for instance, Boolos (1984).

[5] Again we leave the completely analogous formalization of Argument 2 to the reader.

[6] These rules are unobjectionable: they are completely parallel to the usual existential generalization and instantiation rules of first-order logic.

[7] We will later see that comprehension *does* play a role in dealing with Argument 3.

[8] Otherwise, as an anonymous referee rightly observed, there would be no need to appeal to the concept of *arbitrary* finite conjunction in the formalization of Argument 3.

Nonetheless, as we shall now argue, we do not need to appeal to a non-conservative extension of $\mathsf{CT}^-[\mathsf{PA}]$ to derive its conclusion from its premisses.

First, we show how being a finite conjunction can be defined in a natural way in second-order logic. We work in the language of relational second-order logic over the language of arithmetic. In addition to the theorems of PA, we assume that (monadic and relational) second-order comprehension holds for all formulae in the language: we can thus comprehend formulae involving arbitrary arithmetical and relational second-order resources.[9] On the other hand, the *second-order* induction axiom is not assumed in our second-order framework.

We start by being precise about what we will mean by 'finite' in what follows:

> *Definition* 5. We say that a set X is *finite* if it is finitely enumerable. More precisely, we say that X is finite when there is a well-order R on X which is reverse well-founded. (Equivalently, there is a well-order R on X such that: X has an R-last element, and every element of X is either the R-least element or an R-immediate successor of some other element).
>
> *Lemma* 1. If X is finite, then $<$ is a reverse well-founded well-order on X.

Proof. Let R witness the fact that X is finite. Trivially, $<$ is a linear order on X. So, suppose it is not well-founded and let $Y \subseteq X$ have no $<$-least element. We can then define, by recursion on R, a functional relation R' such that (i) R''s domain is X, (ii) if x is the R-least element of X, then $R'(x, y)$ where $y \in Y$ is some arbitrarily chosen object, and (iii) if x is the immediate R-successor of y and $R'(y, z)$, then $R'(x, w)$ where w is the R-least element of Y $<$-below z. If x is the R-greatest element of X and $R'(x, y)$, then there is an element of Y $<$-below y and therefore not in the range of R'. So R' codes a one–one function from X into one of its proper subconcepts, which is impossible.[10] The argument is similar if $<$ is not reverse well-founded. $\square$

Next, we define the way in which a conjunction of a finite set of formulae is inductively built up:

---

[9] Alternatively, we could work in a monadic second-order logic and code relations as concepts of arithmetically coded ordered pairs.

[10] This is so because finite enumerability implies Dedekind-finiteness even in the absence of a Global Well-Ordering Principle.

> *Definition* 6. Say that R is a (canonical) *conjunction sequence* for a finite set of formulae X if (i) R is functional, (ii) R's domain is X, (iii) if x is the $<$-least member of X, then $R(x, x)$, and (iv) if $x \in X$ is the $<$-immediate successor in X of $y \in X$ and $R(y, z)$, then $R(x, w)$ where $w = z \wedge x$. (Since X is assumed to be a set of formulae, $z \wedge x$ is well-defined for $z, x \in X$.)

Next, it can be shown that all finite sets of formulae have conjunction sequences:

> *Lemma* 2. If X is a finite set of formulae, then it has a unique (up to extension) conjunction sequence.

Proof. By Lemma 1, $<$ is a reverse well-founded well-order on X. We prove the existence of unique conjunction sequences by induction on $<$ over X.[11] Clearly, there is such a sequence for the $<$-least element of X and its $<$-predecessors in X. So, suppose R is a conjunction sequence for $x \in X$ and its $<$-predecessors in X. Let $y \in X$ be x's immediate $<$-successor in X, and let z be such that $R(x, z)$. Then $R' = R \cup \langle y, z \wedge y \rangle$ is a conjunction sequence for y and its $<$-predecessors in X. Moreover, since R is unique up to extension, so too is R'. □

> *Definition* 7. When X is a finite set of formulae, let $CONJ(X, x)$ abbreviate the claim that any (equivalently: some) conjunction sequence R for X is such that $R(y, x)$, where y is the $<$-greatest element of X. Let $FIN(X)$ abbreviate the claim that X is a finite set of formulae.

> *Theorem* 8. $\forall X(FIN(X) \rightarrow \exists! x\ CONJ(X, x))$.

Proof. Trivial from Lemma 2. □

If we had an axiom of Global Well-Ordering, we could use Dedekind-finiteness as our notion of finiteness. In the absence of such an axiom, enumerable finiteness (Definition 5) is often taken to be the right notion. However, our argument above with the enumeration notion of finiteness might be regarded as preferable to the strategy with Dedekind-finiteness plus Global Well-Ordering instead. This is because some may see Global Well-Ordering not as a logical but as a mathematical principle, and would then argue that through assuming a Global Well-Ordering, non-conservativeness enters through the back door.

---

[11] Notice that we're not doing induction on $<$ in general, but only on $<$ restricted to X, so we do not need arithmetical induction on second-order formulae to carry it out. We do, however, use arithmetical facts about formulae in the induction, for example, that $x \wedge y$ is well-defined for formulae x and y.

To some extent, our argument is in the spirit of recent projects that attempt to secure conservativeness by separating syntax from subject matter (see, for example, Leigh and Nicolai 2013). However, flat-footedly doing that in response to Fujimoto's argument would be ineffective. Finiteness is, arguably, more an arithmetical property than a syntactic one. In contrast, our notion of finiteness is as natural as the arithmetical one.

It follows from Theorem 8 that we can treat *CONJ* as a function symbol: with mild abuse of language, for any set A, let $CONJ(A)$ be the conjunction of the elements of A if A is a finite set of sentences of $\mathcal{L}_{\mathsf{PA}}$ (and a number that is not the code of a sentence otherwise). Now we can formalize Argument 3 as follows:

(1) $\forall x : Ax\mathsf{PA}(x) \rightarrow Tx$
(2) $\exists X : FIN(X) \wedge \forall y : N(y) \leftrightarrow (y \in X \wedge Ax\mathsf{PA}(y))$
(3) $T(CONJ(N)) \rightarrow Ta$
(4) $T(a)$

In order for the conclusion to be provable from the premisses, we need the following second-order version of conjunctive correctness:

Axiom 2 (CC^2).
$\forall N : FIN(N) \rightarrow (\forall y(Ny \rightarrow Ty) \leftrightarrow T(CONJ(N))$

The principle CC^2 is a very natural way of expressing that an arbitrary finite conjunction is true if and only if all its conjuncts are true. We now show that CC^2 can be conservatively added to CT^-. Let *SOL* be any reasonable system of second-order logic. It may contain full or only restricted second-order comprehension, and second-order choice principles. Then we have:

Proposition 1. The theory $\mathsf{CT}^-[\mathsf{PA}] + SOL + \mathrm{CC}^2$ is first-order arithmetically conservative over PA.

Proof. We show that any model of $\mathsf{CT}^-[\mathsf{PA}]$ can be expanded to a model of $\mathsf{CT}^-[\mathsf{PA}] + SOL + \mathrm{CC}^2$. Since $\mathsf{CT}^-[\mathsf{PA}]$ is first-order arithmetically conservative over PA, this establishes the conclusion.

Take any first-order model $\mathcal{M}$ such that $\mathcal{M} \models \mathsf{CT}^-[\mathsf{PA}]$. Take the *standard* second-order expansion $\mathcal{M}'$ of $\mathcal{M}$ to the language of second-order arithmetic, that is, $\mathcal{M}'$ is like $\mathcal{M}$ except that it also interprets second-order quantifiers, and it takes the second-order quantifiers to range over *all* subsets of the (first-order) domain of $\mathcal{M}$ (standard or nonstandard). We will show that $\mathcal{M}'$ is the expansion we are looking for.

Because $\mathcal{M}'$ interprets the second-order quantifiers in a standard way, it makes *SOL* true. So it suffices to verify that $\mathcal{M}'$ also makes CC^2 true. Again because it is standard for the second-order quantifiers, we have:

$$\mathcal{M}' \models FIN(X) \Leftrightarrow X \text{ is a finite subset of the domain of } \mathcal{M}'.$$

Now (speaking somewhat informally), take any X such that $\mathcal{M}' \models FIN(X)$. Then X really is finite: say it consists of n elements $y_1, \ldots, y_n$ of the domain.

(a) Suppose that for every $i < n$, we have $\mathcal{M}' \models T(y_i)$. We know that $\mathcal{M}' \models \mathsf{CT}^-$, so (by a simple inductive argument in the metalanguage) we see that $\mathcal{M}'$ satisfies

$$(T(y_1) \wedge \ldots \wedge T(y_n)) \rightarrow T(y_1 \wedge \ldots \wedge y_n),$$

and moreover we have $\mathcal{M}' \models CONJ(N) = y_1 \wedge \ldots \wedge y_n$. So we have $\mathcal{M}' \models T(CONJ(N))$.

(b) Conversely, we see in a similar way that if $\mathcal{M}' \models T(CONJ(N))$, then $\mathcal{M}' \models T(y_i)$ for each $i < n$. □

5. Concluding remarks

Field has argued that instances of induction that contain the truth predicate do not count as truth laws because mathematical induction is a mathematical property: it holds in virtue of the natural numbers rather than in virtue of truth.

Finiteness is also a mathematical property. We should be given the freedom to formalize it in one of several acceptable ways. It is not the business of truth theory to prescribe how it should formally be expressed. In particular, *as far as truth theory goes*, it is permissible to treat it as a second-order concept. But we have seen that if we do this, then Fujimoto's non-conservativeness argument no longer goes through. This can be seen as an indication that in formalizing finite correctness as Fujimoto does, using numerical codes of finite sets, one is injecting new *mathematical* content into the theory. Or, in Fieldian terms, the worry is that the principle CC may not be a purely *truth-theoretical principle* after all.

In extending $\mathsf{CT}^-[\mathsf{PA}]$ to a second-order theory, we did not extend the first-order induction scheme of PA to the second-order mathematical induction *axiom*. Someone might object, however, that we *should* do this, and observe that this will result in a second-order theory that is *not* conservative over PA. However, this would amount to conceiving of mathematical induction in an open-ended way. This attitude is, as we have seen, exactly what Field (1999) criticized in his rejection of the 'old' conservativeness argument. In other words, if induction is open-ended, then no *new* conservativeness argument is needed.

There is an interesting remaining question concerning extending $\mathsf{CT}^-[\mathsf{PA}]$ to a second-order setting, however. It might be argued that it is natural to add a compositional truth clause which says that truth also commutes with the second-order quantifiers. Moreover, since not all values of second-order variables have names in the language, it is then natural to switch to satisfaction clauses instead of truth clauses. Let the resulting theory be called $\mathsf{CT}^{2-}[\mathsf{PA}]$. Then the reasoning of the proof of Proposition 2 cannot be used to establish that the extension $\mathsf{CT}^{2-}[\mathsf{PA}] + SOL + \mathrm{CC}^2$ is first-order arithmetically conservative over PA. Whether the resulting theory is conservative or not appears to be an open problem.[12]

References

Boolos, George 1984: 'To Be Is to Be a Value of a Variable (or to Be Some Values of Some Variables)'. *Journal of Philosophy*, 81(8), pp. 430–49.

Enayat, Ali, and Fedor Pakhomov 2019: 'Truth, Disjunction, and Induction'. *Archive for Mathematical Logic*, 58(5–6), pp. 753–66.

Field, Hartry 1999: 'Deflating the Conservativeness Argument'. *Journal of Philosophy*, 96(10), pp. 533–40.

Fujimoto, Kentaro 2012: 'Classes and Truths in Set Theory'. *Annals of Pure and Applied Logic*, 163(11), pp. 1484–1523.

—— 2022: 'The Function of Truth and the Conservativeness Argument'. *Mind*, 131(521), pp. 129–57.

Horsten, Leon, and Graham E. Leigh 2017: 'Truth Is Simple'. *Mind*, 126(501), pp. 195–232.

Leigh, Graham E., and Carlo Nicolai 2013: 'Axiomatic Truth, Syntax and Metatheoretic Reasoning'. *Review of Symbolic Logic*, 6(4), pp. 613–26.

Murzi, Julien, and Lorenzo Rossi 2020: 'Conservative Deflationism?' *Philosophical Studies*, 177(2), pp. 535–49.

---

[12] The authors wish to thank two anonymous referees for extremely helpful comments and suggestions for improvement. The first author's research on this article was made possible by a DAAD-CSC scholarship for post-doctoral research in Germany (University of Konstanz, 2022–24).

Waxman, Daniel 2017: ‘Deflationism, Arithmetic, and the Argument from Conservativeness’. *Mind*, 126(502), pp. 429–63.

Wcisło, Bartosz, and Mateusz Łełyk, M. 2017: ‘Notes on Bounded Induction for the Compositional Truth Predicate’. *Review of Symbolic Logic*, 10(3), pp. 455–80.