Notre Dame Journal of Formal Logic
Volume 59, Number 3, 2018

On the Status of Reflection and Conservativity in Replacement Theories of Truth

Jeffrey R. Schatz

Abstract This article examines Kevin Scharp's formal solution to the alethic paradoxes, ADT, which stands for ascending and descending truth. One of the main supposed benefits of ADT over its competitors is that it alone can validate the uses of truth concepts in theoretical contexts, such as truth-theoretic semantics. The appendixes contain a new consistency proof for ADT, and additionally show that it is conservative. As a result of its conservativity, the article argues that ADT faces a problem in accounting for certain mathematical uses of truth. Thus, Scharp's theory needs to be amended in order to fulfill its aim of replicating all substantive uses of truth.

1 Introduction

Work on the philosophy of truth in the analytic tradition has often centered on the analysis and solution of the alethic paradoxes, especially the liar paradox. A defining feature of the contemporary literature on the liar paradox is the ubiquity of *revenge phenomena*, understood as the tendency for any proposed solution to the liar paradox to generate further paradoxes. As a result, a successful theory of truth must not only solve the liar paradox, but also explain the tendency of further revenge paradoxes to arise in any such theory.

One recent approach to explaining the prevalence of revenge paradoxes is to propose that truth is an inherently problematic concept. In particular, inconsistency theories of truth argue that the constitutive rules governing the use of the concept of truth force an agent to endorse inconsistent claims. As a result, these views claim that revenge paradoxes naturally arise from any solution to the paradoxes that retains each of the meaning-constitutive rules of truth. While some inconsistency theorists, such

Received February 2, 2016; accepted May 11, 2016
First published online June 26, 2018
2010 Mathematics Subject Classification: Primary 03Axx
Keywords: philosophical logic, formal theories of truth, replacement theories, truth, liar paradox
10.1215/00294527-2018-0004

as Matti Eklund and Kirk Ludwig, argue that the inconsistent concept of truth can nonetheless be fruitfully applied in theoretical contexts, other inconsistency theorists argue that the concept of truth must be rejected from any serious semantic theory.[1] This latter approach gives rise to replacement theories of truth, which argue that the truth predicate must be replaced by some new, consistent concepts.

The most prominent example of a replacement theory of truth in the literature is found in Scharp's *Replacing Truth*, which presents a comprehensive theory of truth, involving a formal axiomatization, an explanation of the cause of the alethic paradoxes, and a conception of the nature of truth. For this reason, this article will focus on Scharp's theory as a paradigmatic example of a replacement theory of truth.[2] The theory centers on Scharp's claim that due to the inconsistency of its constitutive principles the truth predicate must, for certain theoretical purposes, be replaced by two new concepts: ascending and descending truth (the theory of these two concepts is henceforth abbreviated ADT). Scharp argues that these two concepts are uniquely able to avoid the alethic paradoxes, retain classical logic, refrain from "monster-barring," and do not require a substantial revision of the theoretical uses of truth. As no other theory of truth can achieve these goals, Scharp argues that ADT provides the best response to the alethic paradoxes.

In this article, I will argue that while ADT is a consistent theory of truth, it fails to live up to all of Scharp's criteria of a successful solution. Specifically, it cannot validate certain applications of truth in mathematical practice. Such applications can be used to argue for the consistency of a formal mathematical theory, something that cannot be proven in the theory itself due to incompleteness. In Appendix B of this paper I show that ADT is conservative, and as such cannot possibly lead to such results. This proof relies on a model construction that is a natural modification of the construction of the minimal model of ADT presented in Appendix A. In particular, here, as often is the case, we prove that one theory is conservative over another by showing that any model of the latter can be expanded to a model of the former. Furthermore, I then show that ADT cannot be supplemented with genuine reflection principles. Then I examine the traditional arguments for why the formulation of such reflection principles are necessary for a formal theory of truth, finding that they hold even in the setting of a replacement theory of truth.

In so doing, this article serves to connect the literature on conservativity and reflection principles with the literature on inconsistency approaches to the alethic paradoxes.[3] The conservativity debate centers around whether a minimal or light concept of truth can and should go beyond the proof-theoretic strength of the original theory. A virtue of inconsistency approaches to the alethic paradoxes is that they can typically explain an increase in the strength of a theory arising from the addition of a truth predicate.[4] As a replacement version of an inconsistency approach, however, ADT does not even permit such an increase in proof-theoretic strength. This article therefore serves as a preliminary exploration of the options for replacement theories in light of this issue.

2 The Theory of Ascending and Descending Truth

Scharp's theory marks a stark distinction between uses of truth in everyday and theoretical contexts (see [13, p. 19]). While truth is used as a device of endorsement and rejection in everyday conversation, these uses of truth rarely generate paradoxes, and, even when they do so, the paradoxes pose little risk to one's purposes and projects. In contrast, the paradoxes can pose far greater problems when they arise in theoretical contexts where one employs truth in systematic investigations of meaning and reasoning (see [13, p. 1]). Scharp suggests that, as an inconsistent concept, truth should be replaced specifically for these theoretical purposes with a new concept: descending truth. In this article, I will focus solely on Scharp's axiomatization of descending truth.

The property of semantic release—the ability to infer φ from $D(\varphi)$—is constitutive of descending truth, as this property enables descending truth to function as a device of endorsement. Beyond semantic release, descending truth is wholly characterized by seven axioms (the theory ADT):[5]

Axiom D1: $D(\ulcorner\varphi\urcorner) \to \varphi$; if φ is an $L[D]$-sentence
Axiom D2: $D(\ulcorner\neg\varphi\urcorner) \to \neg D(\ulcorner\varphi\urcorner)$
Axiom D3: $D(\ulcorner\varphi \wedge \psi\urcorner) \to (D(\ulcorner\varphi\urcorner) \wedge D(\ulcorner\psi\urcorner))$
Axiom D4: $[D(\ulcorner\varphi\urcorner) \vee D(\ulcorner\psi\urcorner)] \to D(\ulcorner\varphi \vee \psi\urcorner)$
Axiom D5: $D(\ulcorner\varphi\urcorner)$; if φ is a logical truth
Axiom D6: $D(\ulcorner\varphi\urcorner)$; if φ is a theorem of PA with the induction schema for the language of PA supplemented with a D predicate[6]
Axiom D7: $D(\ulcorner\varphi\urcorner)$; if φ is an instance of D1–D6

Throughout this article, an $L[D]$-sentence is defined as any sentence, understood in the typical way, in the language of Peano arithmetic with an additional descending truth predicate D. Additionally, a logical truth will be understood as meaning that a formula has a deduction in first-order logic. We will understand $\ulcorner\varphi\urcorner$ to be the standard Gödel numbering for the formula φ, though to ease readability we will sometimes drop this notation when convenient. We note that ADT is in fact a consistent theory, relative to the consistency of Peano arithmetic augmented with a single predicate D.[7] (For a consistency proof, which additionally identifies the minimal ω-model of ADT, see Appendix A.)

Given these axioms, Scharp also defines two further concepts. Ascending truth is defined for all sentences φ as $\neg D(\neg\varphi)$; we will abbreviate that φ is ascending true as $A(\varphi)$ (see [13, p. 152]). Semantic capture—the ability to infer $A(\varphi)$ from φ—is constitutive of ascending truth's meaning. Semantic capture enables ascending truth to serve as a device of rejection, as $\neg\varphi$ follows from $\neg A(\varphi)$ (see [13, p. 149]). Further, safety is defined for all sentences φ as $A(\varphi) \to D(\varphi)$. If some sentence φ is safe, then descending truth and ascending truth can be replaced by the naive concept of truth without fear of paradox (see [13, p. 153]).

Scharp argues that these principles are able to serve as a satisfactory replacement for the inconsistent natural language truth predicate. First of all, descending truth validates the rule (T-Out) ($T(\varphi) \to \varphi$) while ascending truth validates (T-In) ($\varphi \to T(\varphi)$). As these rules are at least some of the constitutive principles determining the meaning of truth, any replacement for truth must be able to validate them in some form: "the truth predicates play several important roles in our linguistic practices and these roles depend on the truth predicates obeying (T-In) and (T-Out)" [13, p. 63].[8] Furthermore, Scharp claims that ADT, alone among possible solutions to the alethic paradoxes, is able to validate all of the established theoretical uses of truth.

As a result, ADT fundamentally gets the data right regarding the natural language uses of the concept of “truth.”

Scharp pays particularly close attention to the uses of truth in linguistics. Scharp rejects consistency theories of truth because such theories do not enable truth-theoretic semantics, a cornerstone of contemporary semantics: “Unless truth is treated as an inconsistent concept, we have no hope for a truth-conditional semantics for natural language” [13, p. 121]. For Scharp, to question such successful uses of truth by various branches of science is tantamount to a dereliction of one’s duty as a philosopher. Philosophy serves to generate conceptual schemes acceptable for theoretical usage, but not to reject these uses for purely philosophical reasons. One’s philosophical account of truth must follow one’s prior uses of the truth predicate, and not the other way around: “If one’s favored view of truth . . . conflicts with the sciences, then it is the philosophical view that should go” [13, p. 123].

3 The Problem of Conservativity in ADT

Scharp’s strong commitment to validating theoretical uses of the truth predicate poses a significant challenge for ADT in relation to the problem of conservativeness. In [16], Shapiro raises the question of whether a theory of truth should be conservative, that is, whether the extension of a theory in some original language to a language able to express a truth predicate should be a conservative extension of the original theory. Shapiro [16, p. 499] argues that a theory of truth, at least when supplemented by mathematical induction for predicates in the language of the truth predicate, should be nonconservative, as such an expanded induction schema can be used to inductively prove the consistency of the base mathematical theory. As a result of Gödel’s second incompleteness theorem, this fact could not be proved in the original theory, provided that it was in fact consistent. Therefore, if a theory of truth is able to prove such a consistency statement, then it must be nonconservative. As Shapiro holds that an expanded induction schema is necessary for any serious candidate for a formal explication of truth, he thus concludes that any satisfactory candidate for a formal theory of truth must be nonconservative (see [16, p. 500]).

Let us spell a particular mathematical use of truth out further: arguing for the consistency of ZFC. By noting both the truth of the axioms of ZFC and that logically valid inferences preserve truth, any arbitrary consequence of these axioms can be endorsed. Due to the generality of this procedure, one can conclude that any consequence must be true, and therefore that ZFC is consistent. Such argumentation has been endorsed by a wide spectrum of philosophers of mathematics, often in response to the second incompleteness theorem. Despite the independence of any strong mathematical theory’s Gödel sentence or consistency statement, such reasoning supposedly establishes the possibility of inferring these claims from the mere acceptance of the strong theory itself.[9] Due to the prevalence of such truth-based arguments in the literature on the second incompleteness theorem, there is prima facie reason to believe that these uses of truth in mathematics are well established in a similar manner as truth-theoretic semantics is in linguistics.

As a result, the uses of truth predicates in mathematical argumentation seem as established as the linguistic uses which Scharp argues we cannot reject. Yet, this possibility poses a nontrivial challenge for Scharp’s system, as it can be demonstrated that ADT is a conservative extension of certain strong mathematical theories, such as ZFC (see Appendix B). It follows that the addition of Scharp's replacement concepts to some mathematical theories does not permit the proof of new results in the language of ZFC. As such results, including the Gödel sentence of ZFC and the statement of the consistency of ZFC, are the immediate consequences of the use of truth in mathematical reasoning, ADT seemingly cannot capture this use of truth. This conclusion would imply that Scharp is forced to be revisionary toward certain elements of mathematical practice in a parallel way to how alternative theories of truth require revisions of linguistic practice. Scharp explicitly requires that any response to the alethic paradoxes be able to capture all of the accepted theoretical uses of truth. Yet, if one accepts the uses of truth in arguing for independent statements of strong mathematical theories, then there is a seemingly substantive mathematical use of truth that escapes ADT. Thus, his approach would fail to meet the strong requirements that he himself sets forth.

There are two defenses against this challenge for Scharp's theory. The first defense would claim that, despite the conservativity of ADT, the theory can nonetheless somehow capture the mathematical uses of truth. Such a response would dissolve the above challenge, showing that the essential mathematical uses of truth are in fact separate from the conservativity issue. The alternative defense would be to recognize the failure of ADT to permit the supposed uses of the concept of truth in mathematics, but claim that the traditional arguments for the importance of such uses tacitly assume that a theory is seeking to retain the very alethic concept which Scharp seeks to replace. As a result, the lack of reflection principles in ADT would fail to represent a serious challenge for a replacement theory of truth. We will consider the merits of these defenses in the following two sections.

4 Reflection Principles and Mathematical Uses of Truth

In considering important features of the uses of truth in mathematical reasoning, we will focus on the case of proving the consistency of ZFC. Here we seek to connect the notions of provability in a certain formal system with assertability or mathematical truth: "There is also the intuitive principle that if a proposition or sentence is proven, then it is true" [13, p. 205]. It is often claimed that the intuitive notion of proof requires the assumption of such a principle to justify the use of proofs. Otherwise, one would have no reason to believe that a proof would provide support for derived conclusions. Any theory of truth that was capable of replicating this reasoning would need to have a counterpart for these features.

Such an inferential role of semantic reasoning can be formalized through a formal reflection principle. A reflection principle is a proposition in a mathematical language that in some manner internalizes the claim that a certain mathematical theory is consistent. When a mathematical theory is supplemented with a truth predicate, these principles often take the form "For all φ, if φ is proven in S, then φ is true" for the relevant theory S.[10] As a result of the connections between reflection principles and the consistency of a theory, incompleteness shows that we cannot include a theory's own reflection principle in it under pain of a contradiction: "a reflection principle ... states something about a formal theory that cannot be captured by the formal theory on pain of contradiction" [13, p. 118]. In particular, the reflection principle instance "$\mathrm{Prv}_{\mathrm{ZFC}}(0 = 1) \rightarrow T(0 = 1)$" together with "$\neg T(0 = 1)$" entails

"$\neg\mathrm{Prv}_{\mathrm{ZFC}}(0 = 1)$." Provided that the base theory ZFC is consistent, and therefore does not prove every statement, "$\neg\mathrm{Prv}_{\mathrm{ZFC}}(0 = 1)$" can be seen as a consistency statement by expressing that there is in fact some particular statement that ZFC fails to prove.

It is important to note that the mere expressibility of a reflection principle does not by itself entail the nonconservativity of a theory of truth. In the case of an individual proposition, the reflection principle merely entails that some provable proposition is true. By itself, such an entailment does not imply anything in the language of the base mathematical theory that was not implied by the mathematical theory itself. Furthermore, the proof of the "$\neg\mathrm{Prv}_{\mathrm{ZFC}}(0 = 1)$" through a reflection principle employs an auxiliary hypothesis—namely, that the theory can derive the falsity of some contradiction, such as "$\neg T(0 = 1)$."[11] As this further hypothesis may not be derivable in a formal theory of truth, such instances of reflection principles can be independent of the question of conservativeness. As a result, merely having expressible reflection principles in ADT need not conflict with the conservativity of the theory.

Due to the close connections between reflection principles and the inferential uses of truth in mathematical arguments, the advocate of ADT could argue that the theory can capture these mathematical uses of truth if one could derive reflection principles in the theory. Such an argument would be in keeping with the first possible response to the challenge of the conservativity of ADT. For a theory that retains the concept of truth, there is only one option for a reflection principle, but for a replacement theorist like Scharp, there are multiple options for formulating a reflection principle. As Scharp replaces truth with two distinct concepts, there is a possibility of two reflection principles, Reflection-D and Reflection-A:[12]

Reflection-D: If a sentence is proven, then it is descending true.
Reflection-A: If a sentence is proven, then it is ascending true.

Of these two possibilities, the formulation Reflection-D is certainly able to fulfill the role of a reflection principle. Since ADT entails "$\neg D(0 = 1)$," Reflection-D can be used to derive a formal consistency statement for strong mathematical theories like ZFC. Thus, if Reflection-D were a consequence of ADT, then ADT would succeed in validating these inferences, conservativity notwithstanding. Nonetheless, Scharp notes that Reflection-D is inconsistent with ADT (see [13, p. 205]). Consider the descending liar λ, which states that λ is not descending true. We note that ADT proves that λ is not descending true. As $D(\lambda) \rightarrow \lambda$, if λ were descending true, then it would follow that λ. But this is equivalent to $\neg D(\lambda)$, generating a contradiction. Thus, we can derive $\neg D(\lambda)$. As this is equivalent to the content of λ, ADT proves λ. Thus, Reflection-D would entail that λ was descending true, leading to it being both descending true and not descending true, a clear contradiction. As a result, Reflection-D cannot successfully fulfill the role of a reflection principle in ADT.

On the other hand, Reflection-A is consistent with ADT, since if ADT proves φ, then, by the constitutive principle of ascending truth, φ is ascending true on any model of ADT. Scharp takes this as providing a conceptual explication of proof in ADT: "We should use Reflection-A as our conceptual connection between proof and the replacement concepts" [13, p. 205]. Additionally, Reflection-A has the syntactic form typical of a reflection principle. As a result, one might expect Reflection-A to function as the equivalent of a reflection principle for ADT. Nonetheless, this is not the case. Reflection-A can only be used to derive conclusions of the form that φ is ascending true, which provides neither an endorsement nor a rejection of φ. As statements of the form $A(\varphi)$ can entail neither φ nor its negation, Reflection-A wholly fails to be useful in permitting any justification or knowledge-generating inferences.[13] Thus, Reflection-A cannot be used in the inferences characteristic of a genuine reflection principle. I posit that a replacement for some principle must be capable of fulfilling the original principle's important inferential roles, yet the replacement reflection principle cannot do so. For this reason, Scharp seems mistaken to ascribe such significance to Reflection-A.

It may be instructive to identify precisely where the standard inductive proof through reflection principles to Con(PA) fails when ascending truth is used in place of truth in formulating the reflection principles.[14] The standard proof begins by claiming that a formal theory of truth for a base theory S proves that all axioms of S are true. The standard proof then claims that logically valid inferences are truth preserving; that is, if the premises of a valid inference are true, then the conclusion of the inference must also be true. It is at this step, however, that the argument fails when truth is replaced with ascending truth. Scharp notes that, for both ascending and descending truth, "one should not accept that valid arguments are necessarily truth-preserving" [13, p. 151]. As a result, a crucial step in the argument fails, and so the standard formal proof through reflection principles fails to entail the consistency of PA when truth is replaced with ascending truth.

Thus, there are two candidates for a reflection principle for ADT.[15] The first, Reflection-D, would function as a genuine reflection principle, but is unable to consistently be added to ADT. The second, Reflection-A, while compatible with ADT, cannot actually be used as a reflection principle. Thus, neither of these two options are up to the task of enabling the mathematical inferences associated with naive semantic reasoning. The question arises of whether there is some other principle that could be both consistently added to ADT and be powerful enough to fulfill the role of a reflection principle. This prospect seems dim, as it is unclear both of what syntactic form it could be and how it could be strong enough to lead to nonconservative inferences. Due to the close connection between reflection principles and the mathematical uses of truth, as well as the established conservativity of the theory, it therefore seems that ADT cannot in fact capture these uses of truth. Thus, the first potential response to our challenge regarding the conservativity of ADT seems to fail.

5 Current Criticisms of Reflection Principles

A second possible response to the challenge arising from the conservativity of ADT would be to concede the inability of ADT to replicate the supposed mathematical uses of truth, but argue that this poses no problem for a replacement theory of truth. The traditional arguments for the importance of the concept of truth in mathematical argumentation, as put forth by Mostowski and Shapiro, have recently come under considerable scrutiny, with various philosophers challenging whether any acceptable theory of truth must endorse them.[16] As Walter Dean [1] understands it, the traditional arguments center around the implicit commitment thesis (henceforth ICT), which claims that anyone who accepts a mathematical theory is already committed to accepting further statements—including reflection principles—that are formally independent of the theory (see [1, pp. 32–34]). These commitments arise by noting that if all of the axioms of a theory are true and if all the inferences used in the theory are truth preserving, then all of the consequences of the theory itself must be true. The ICT argument posits that, while these reflection principles cannot even be formulated in the original language of the mathematical theory itself, they are already wholly justified by a subject's epistemic attitudes to the original theory. Given that reflection principles are already justified, advocates of ICT argue that any adequate theory of truth should—solely by virtue of expanding the language of the original theory—permit the derivation of these principles.

Dean argues that there are two distinct problems with the ICT argument, potentially providing a rationale for a formal theory of truth that is unable to articulate reflection principles. The first problem concerns the scope of applicability for the ICT argument. While advocates of ICT claim that the argument forces any mathematical theorist to accept the reflection principles for any endorsed theory, Dean notes that formal reflection principles in various forms correspond directly to strong versions of mathematical induction. As a result, it may be consistent for proponents of theories which withhold acceptance of the full schema of mathematical induction—including forms of finitism and predicativism—to remain agnostic toward the reflection principles for such theories (see [1, p. 54]). By this argument, one could avoid a commitment to reflection principles by refraining from accepting certain mathematical claims, including the full principle of mathematical induction for PA. Given Scharp's commitment to not revising prior theoretical uses of truth, however, it would be improper to choose one's mathematical commitments for the sake of defending ADT. Doing so would amount to favoring a philosophical theory of truth over mathematical practice in precisely the way that Scharp condemns. Thus, Dean's first problem does nothing to shield ADT from the impact of the ICT argument.

Furthermore, Dean also considers whether the ICT argument compels even the classical mathematical theorist to endorse reflection principles. Considering the relationship between the informal argument for the consistency of PA and more formal derivations of this claim, Dean notes they are separated by a genuine mathematical principle—namely, the extension of the induction schema to include sentences involving the truth predicate (see [1, p. 61]). In particular, the formal argument for reflection principles relies on an induction on the length of proofs showing that all theorems of PA have the property of truth. As a result, Dean concludes that the ICT requires the derivation of reflection principles from a formal theory of truth only given the assumption that the truth predicate can be fully utilized in the induction schema. ADT then would be justified in not proving the descending reflection principle if the theory also limited the use of descending truth in the induction schema of PA. Though Scharp does not speak explicitly on the scope of the induction schema in ADT, there seems little prospect for such a response given Scharp's further commitments.[17] In particular, Scharp sees PA as the theory of syntax underlying ADT, and on that basis finds that any attempt to limit the resources of PA in ADT is wholly improper (see [13, p. 152]). As the induction schema is among the axioms of PA, Scharp's explicit commitments seem to require an acceptance of the full induction schema in the language of ADT. Thus, both problems Dean notes for the capability of the ICT argument to require any formal theory of truth to derive formal reflection rely on a theorist rejecting genuine mathematical principles, but in both cases Scharp seems committed to accepting these principles.

Separately from Dean's critique of the ICT argument, Harty Field [4] has also challenged the claim that any satisfactory formal theory of truth must derive certain reflection principles. Field notes that the typical informal argument for the consistency of a formal theory relies on two essential premises: that all axioms of the theory are true and that all inferences in the theory preserve truth (see [4, p. 201]). These assumptions cannot be accepted for the entirety of a formal theory of truth, however, without leading to a contradiction on the basis of Gödel's second incompleteness theorem. As a result, Field concludes that any serious and noncontradictory theory of truth must reject the claim that all accepted logical inferences are truth preserving (see [4, p. 203]). Without the claim that all first-order inferences are truth preserving, the use of truth to argue informally for the consistency of even a subtheory of the formal theory of truth—such as PA or ZFC—fails, and so Field argues that both the consistency statements and reflection principles of such theories need not be derivable in a satisfactory formal theory of truth.

At first, this argument seems to provide a solid defense for Scharp regarding the underivability of reflection principles in ADT. Scharp explicitly notes that these issues necessarily lead to the failure of ADT to prove that all logical inferences preserve either ascending or descending truth (see [13, p. 151]). However, as Scharp intends for ADT to be interpreted, all sentences in the language of PA should be either both ascending and descending true or both ascending and descending false.[18] Additionally, ADT can derive the restricted claim that all rules of inference, as applied to only such sentences, are both ascending and descending truth preserving.[19] Thus, Scharp seems committed to the claim that the rules of inference are truth preserving when applied to sentences in the language of PA. As a result, Field's argument only provides a defense for the inability of ADT to prove its own reflection principle, and not a defense of its inability to prove the reflection principles for PA and ZFC. Thus, the informal argument for the importance of reflection principles still seems to apply to ADT.

As a result, we find that the usual considerations suggestive of the significance of deriving reflection principles for accepted mathematical theories in a satisfactory formal theory of truth stands even for a replacement theory of truth like ADT. Through Scharp's strong commitment to respecting theoretical practice with respect to the truth predicate, he accepts the premises underlying such arguments. It follows that there is little prospect for Scharp consistently rejecting the importance of reflection principles for ADT. Thus, the second defensive strategy available to Scharp—rejecting the arguments underlying reflection principles—is of no use in dissolving the problems of reflection principles. Given that there is also little prospect for ADT to endorse or even formulate reflection principles, we find that the conservativity of ADT does pose a genuine challenge to Scharp's goal of respecting mathematical uses of truth.

Given the dim prospects for rejecting the importance of reflection principles—and therefore also the mathematical uses of truth—Scharp seemingly must accept a third response to the problem of conservativeness: that is, accept that this poses a genuine problem for ADT. As it currently stands, ADT fails to live up to its promises of respecting all of the scientific uses of truth. While the theories that Scharp rejects fail to support linguistic practice, ADT instead fails to support mathematical practice. It follows that ADT is not strictly better than other theories according to Scharp's own standards, but instead represents a trade-off, privileging the linguistic uses of the concept of truth over the mathematical uses.

6 Conclusion

Scharp argues that ADT is superior to all of the alternative formal theories of truth due to a variety of criteria, but first and foremost is its ability to validate all established theoretical uses of truth. While he does establish that his theory, seemingly alone among the various alternatives, can validate many of the received uses of truth, this article has shown that it does so at a cost. Unlike its competitors, ADT cannot endorse the mathematical uses of truth. As a result, though this article has established that ADT is a consistent formal theory of truth, and thus a genuine option, it fails to wholly succeed on all of Scharp's criteria for such a theory.

Additionally, we note that this article has served to connect the literature on ADT, the foremost example of a replacement theory on offer in the extant literature, with the traditional literature on conservativity and reflection principles in formal theories of truth. While most of the latter literature implicitly assumes that such a theory seeks to retain and utilize the concept of truth, we have found that the arguments for nonconservativity and the importance of reflection principles nonetheless apply to a replacement theorist as well. Regardless of this fact, we conclude by noting that replacement theories of truth offer a new and interesting approach to the problems arising from the alethic paradoxes, and that much further work remains to be done connecting these new approaches with traditional arguments and issues from the broader literature on truth.

Appendix A: Consistency of ADT

The first question that arises about ADT is whether it is a consistent theory or not.[20] We will prove the consistency of ADT with a three-step construction. In the first step, the minimal ω-model of axioms D2–D6 will be constructed. Next, this model will be shown to be the minimal ω-model for D1–D6 as well. Finally, using this model as a starting point, the minimal ω-model of ADT will be constructed.

Let $L[D]$ be an expansion of the signature L of PA by a new unary predicate symbol D. Then ω-models are written as $(\omega, \mathbb{D})$, where $\mathbb{D} \subseteq \omega$. By abuse of notation, we sometimes refer to $\mathbb{D}$ as an ω-model and write $\mathbb{D} \models \varphi$ in lieu of $(\omega, \mathbb{D}) \models \varphi$, and likewise we write $\mathbb{D} \models T$ in lieu of $(\omega, \mathbb{D}) \models T$. For the remainder of what follows, we fix a recursively enumerable consistent deductively closed $L[D]$-theory T_0 extending PA$[D]$ which is likewise true on all ω-models. Here PA$[D]$ is simply PA with the induction schema for $L[D]$-formulas. For the sake of concreteness, we could take T_0 to be the deductive closure of PA$[D]$. Note that the following proof is for the general case where D6 is replaced with a generalization, D6(T_0), which states that $D(\varphi)$ for all φ such that $T_0 \models \varphi$. Clearly, D6$(T_0) \models$ D6.

Let $T_0^{\sharp}$ be the $L[D]$-theory having D1–D7 as axioms. Instead of saying that φ is an instance of D1–D6(T_0), we might sometimes say that $\varphi \in$ D1–D6. Further, let

$T_0^\dagger$ be $T_0^\sharp$ minus D7, and let $T_0^\flat$ be $T_0^\dagger$ minus D1. If T is any theory in the signature of $L[D]$, then we say that T has a *minimal ω-model* if there is an ω-model $\mathbb{D}_0 \models T$ such that for all ω-models $\mathbb{D} \models T$, we have $\mathbb{D}_0 \subseteq \mathbb{D}$.

In the following propositions, we construct the minimal model of $\mathbb{D}_0^\sharp$ by first constructing the minimal models for $\mathbb{D}_0^\flat$ and $\mathbb{D}_0^\dagger$. We know from axiom D6 that $\mathbb{D}$ in $\mathbb{D}_0^\sharp$ must at least include the deductive closure of PA. One might wonder how much bigger the extension of $\mathbb{D}$ has to be than this. The main idea behind these model constructions is that its extension does not need to be much larger than the deductive closure of PA. In fact, the extension of $\mathbb{D}$ in the minimal model of $\mathbb{D}_0^\sharp$ will consist in the deductive closure of our theory T_0 together with instances of the axioms D1–D6 and finite disjunctions including an instance of D1–D6 as one of the disjuncts.

Proposition A.1 *The minimal ω-model of $T_0^\flat$ is $\mathbb{D}_0^\flat = \{\ulcorner\varphi\urcorner : T_0 \models \varphi\}$. Hence, if $\ulcorner\varphi\urcorner \in \mathbb{D}_0^\flat$, then φ is true on all ω-models of T_0.*

Proof First, we show that $\mathbb{D}_0^\flat \models$ D2. Let $\neg\varphi \in \mathbb{D}_0^\flat$. This implies that $T_0 \models \neg\varphi$. Assume that $\varphi \in \mathbb{D}_0^\flat$. This likewise implies that $T_0 \models \varphi$. But then T_0 is a model of both φ and $\neg\varphi$, and, as such, is inconsistent. But, by assumption, T_0 is consistent. As a result, $\neg(\varphi \in \mathbb{D}_0^\flat)$, and so $\mathbb{D}_0^\flat \models$ D2.

Next, we show that $\mathbb{D}_0^\flat \models$ D3. Let $\varphi \wedge \psi \in \mathbb{D}_0^\flat$. This implies that $T_0 \models \varphi \wedge \psi$. As T_0 is deductively closed, this means that $T_0 \models \varphi$ and $T_0 \models \psi$. As a result, $\varphi \in \mathbb{D}_0^\flat$ and $\psi \in \mathbb{D}_0^\flat$. As a result, $\mathbb{D}_0^\flat \models$ D3.

Next, we show that $\mathbb{D}_0^\flat \models$ D4. Let $\varphi \in \mathbb{D}_0^\flat$ or $\psi \in \mathbb{D}_0^\flat$. In the first case, this means that $T_0 \models \varphi$, and, as T_0 is deductively closed, $T_0 \models \varphi \vee \psi$. As a result, $\varphi \vee \psi \in \mathbb{D}_0^\flat$. Similarly, in the second case, $T_0 \models \psi$, and, by the deductive closure of T_0, $T_0 \models \varphi \vee \psi$. Thus, $\varphi \vee \psi \in \mathbb{D}_0^\flat$. In either case, $\varphi \vee \psi \in \mathbb{D}_0^\flat$. Thus, $\mathbb{D}_0^\flat \models$ D4.

Next, we show that $\mathbb{D}_0^\flat \models$ D5. Assume that φ is a logical truth. By definition, logical truths are modeled by all first-order structures. Thus, $T_0 \models \varphi$. As a result, $\varphi \in \mathbb{D}_0^\flat$. Thus, $\mathbb{D}_0^\flat \models$ D5.

Finally, we show that $\mathbb{D}_0^\flat \models$ D6. Assume that φ is a theorem of PA[D]. As T_0 includes the deductive closure of PA[D], $T_0 \models \varphi$. Thus, $\varphi \in \mathbb{D}_0^\flat$. Therefore, $\mathbb{D}_0^\flat \models$ D6.

Putting the above together, we find that $\mathbb{D}_0^\flat$ is a model of $T_0^\flat$. We must demonstrate that it is the minimal such model. Let $\varphi \in \mathbb{D}_0^\flat$, and let $\mathbb{D}$ be an ω-model of $T_0^\flat$. We know that $T_0 \models \varphi$ by the definition of $\mathbb{D}_0^\flat$. As $\mathbb{D} \models T_0^\flat$ and $T_0^\flat \models \mathrm{D6}(T_0)$, we know that $\mathbb{D} \models D(\varphi)$. Thus, $\varphi \in \mathbb{D}$. By generality, for any ω-model $\mathbb{D}$ of $T_0^\flat$, we find that for any φ such that $\varphi \in \mathbb{D}_0^\flat$, $\varphi \in \mathbb{D}$; that is, $\mathbb{D}_0^\flat \subseteq \mathbb{D}$. Thus, $\mathbb{D}_0^\flat$ is the minimal ω-model of $T_0^\flat$. □

Proposition A.2 *$\mathbb{D}_0^\flat$ is also a model of $T_0^\dagger$, and it is the minimal ω-model of $T_0^\dagger$.*

Proof By Proposition A.1, we know that $\mathbb{D}_0^\flat$ is a model of D2–D6. Hence, we must merely show that it is a model of D1. Assume that $\mathbb{D}_0^\flat \models D(\varphi)$. By definition, this means that $\varphi \in \mathbb{D}_0^\flat$, that is, that $T_0 \models \varphi$. As $\mathbb{D}_0^\flat \models T_0$ by the definition of $\mathbb{D}_0^\flat$, we find that $\mathbb{D}_0^\flat \models \varphi$. Thus, $\mathbb{D}_0^\flat \models D(\varphi) \rightarrow \varphi$. Thus, $\mathbb{D}_0^\flat \models D1$, and we find that $\mathbb{D}_0^\flat \models T_0^\dagger$.

We must further show that $\mathbb{D}_0^\flat$ is the minimal model of $T_0^\dagger$. Let $\varphi \in \mathbb{D}_0^\flat$, and let $\mathbb{D}$ be an ω-model of $T_0^\dagger$. As $T_0^\flat \subseteq T_0^\dagger$, we know that $\mathbb{D} \models T_0^\flat$. By Proposition A.1, we know that $\mathbb{D}_0^\flat$ is the minimal ω-model of $T_0^\flat$, so $\mathbb{D}_0^\flat \subseteq \mathbb{D}$. Thus, $\mathbb{D}_0^\flat$ is the minimal ω-model of $T_0^\dagger$. □

Proposition A.3 *The minimal ω-model of $T_0^\sharp$ is*

$$\begin{aligned}\mathbb{D}_0^\sharp &= \mathbb{D}_0^\flat \cup \{\ulcorner \varphi \urcorner : \varphi \in \text{D1–D6}\} \\ &\quad \cup \left\{\ulcorner \bigvee_{i=1}^{n} \varphi_i \urcorner : \varphi_i \in \text{D1–D6}\, \textit{for some } 1 \leq i \leq n\right\}. \end{aligned} \tag{A.1}$$

Further, if $\ulcorner \varphi \urcorner \in \mathbb{D}_0^\sharp$, *then* φ *is true on* $\mathbb{D}_0^\flat$.

Proof First, we verify that $\mathbb{D}_0^\sharp \models$ D5–D7. By Proposition A.2, since $\mathbb{D}_0^\flat \subset \mathbb{D}_0^\sharp$, we know that $\mathbb{D}_0^\sharp \models$ D5–D6. Similarly, since for any $\varphi \in$ D1–D6 we have that $\varphi \in \mathbb{D}_0^\sharp$, we know that $\mathbb{D}_0^\sharp \models$ D7.

Next, we verify that $\mathbb{D}_0^\sharp \models$ D3. Suppose that $\mathbb{D}_0^\sharp \models D(\ulcorner \varphi \wedge \psi \urcorner)$. Then $\ulcorner \varphi \wedge \psi \urcorner \in \mathbb{D}_0^\sharp$. There are then three cases to consider. First, suppose that $\ulcorner \varphi \wedge \psi \urcorner \in \mathbb{D}_0^\flat$. Then by the deductive closure of T_0 we have $\ulcorner \varphi \urcorner, \ulcorner \psi \urcorner \in \mathbb{D}_0^\flat \subseteq \mathbb{D}_0^\sharp$. Second, suppose that $\varphi \wedge \psi$ is an instance of D1–D6. Since the main connective of $\varphi \wedge \psi$ is a conjunction, we see that this case is in fact impossible by inspection of the logical form of D1–D6. Specifically, when we inspect the axioms D1–D6 we find that they all either are atomic formulas or have a conditional as their main connective, and so we find that no statement with a conjunction as its main connective could be an instance of one of these axioms. And similarly for the case that $\varphi \wedge \psi$ has the form $\bigvee_{i=1}^{n} \varphi_i$, where one of the φ_i is from D1–D6. Thus, we find that $\mathbb{D}_0^\sharp \models \varphi$ and that $\mathbb{D}_0^\sharp \models \psi$ in all three cases. Therefore, we find that $\mathbb{D}_0^\sharp \models$ D3.

Next, we verify that $\mathbb{D}_0^\sharp \models$ D4. Suppose that $\mathbb{D}_0^\sharp \models D(\ulcorner \varphi \urcorner) \vee D(\ulcorner \psi \urcorner)$. Then $\varphi \in \mathbb{D}_0^\sharp$ or $\psi \in \mathbb{D}_0^\sharp$. Without loss of generality, assume that $\varphi \in \mathbb{D}_0^\sharp$. Again, there are three cases to consider. If $\varphi \in \mathbb{D}_0^\flat$, then, by the deductive closure of $\mathbb{D}_0^\flat$, we have $\ulcorner \varphi \vee \psi \urcorner \in \mathbb{D}_0^\flat \subset \mathbb{D}_0^\sharp$. Next, assume that $\varphi \in$ D1–D6. Then $\varphi \vee \psi$ is of the form $\varphi_1 \vee \varphi_2$, where $\varphi_1 \in$ D1–D6. As a result, $\varphi \vee \psi$ is of the form $\ulcorner \bigvee_{i=1}^{n} \varphi_i \urcorner : \varphi_i \in$ D1–D6 for some i such that $1 \leq i \leq n$. Thus, $\varphi \vee \psi \in \mathbb{D}_0^\sharp$ by the third component of its definition. Finally, assume that φ has the form $\bigvee_{i=1}^{n} \varphi_i$, where some $\varphi_i \in$ D1–D6. This implies that $\varphi \vee \psi$ is also of the form $\bigvee_{i=1}^{n} \varphi_i$. Thus, $\varphi \vee \psi \in \mathbb{D}_0^\sharp$. Thus, in all three cases, $\mathbb{D}_0^\sharp \models \varphi \vee \psi$. Therefore, we find that $\mathbb{D}_0^\sharp \models$ D4.

Next, we verify that $\mathbb{D}_0^\sharp \models$ D2. First, we show that if $\varphi \in \mathbb{D}_0^\sharp$, then φ is true on $\mathbb{D}_0^\flat$. Let $\varphi \in \mathbb{D}_0^\sharp$. If $\varphi \in \mathbb{D}_0^\flat$, then, by Proposition A.1, φ is true on all ω-models of $T_0^\flat$. In particular, it is true on $\mathbb{D}_0^\flat$. If $\varphi \in$ D1–D6, then, as $\mathbb{D}_0^\flat \models T_0^\dagger$ by Proposition A.1, we know that φ is true on $\mathbb{D}_0^\flat$. Finally, if φ is a disjunction including some $\psi \in$ D1–D6 as a disjunct, then, as D1–D6 are true on $\mathbb{D}_0^\flat$, φ is also true on $\mathbb{D}_0^\flat$. In all three cases, φ is true on $\mathbb{D}_0^\flat$. Thus, all $\varphi \in \mathbb{D}_0^\sharp$ are true on $\mathbb{D}_0^\flat$.

Let $\ulcorner \neg\varphi \urcorner \in \mathbb{D}_0^\sharp$. As shown above, this implies that $\neg\varphi$ is true on $\mathbb{D}_0^\flat$. Assume that $\ulcorner \varphi \urcorner \in \mathbb{D}_0^\sharp$. Again, by the above this implies that φ is true on $\mathbb{D}_0^\flat$. So both φ and $\neg\varphi$ are true on $\mathbb{D}_0^\flat$. As a result, we find that the assumption that $\ulcorner \varphi \urcorner \in \mathbb{D}_0^\sharp$ leads to a direct contradiction. Thus, we find that $\ulcorner \varphi \urcorner \notin \mathbb{D}_0^\sharp$, and, as such, that $\mathbb{D}_0^\sharp \models$ D2.

Finally, we verify that $\mathbb{D}_0^\sharp \models$ D1. Where $|\varphi|$ indicates the length of formula φ, we will prove by induction on the length of formulas l that $|\varphi| \leq l \rightarrow [\varphi \in \mathbb{D}_0^\sharp \rightarrow \mathbb{D}_0^\sharp \models \varphi]$. As there are no formulas of length 0, clearly this holds for $l = 0$.

Assume that $\varphi \in \mathbb{D}_0^\sharp$, that $|\varphi| = l + 1$, and that the induction hypothesis holds for formulas of length $n \leq l$. First, assume that $\varphi \in \mathbb{D}_0^\flat$. By Proposition A.1, this implies that φ is true on all ω-models of $T_0^\flat$. Thus, $\mathbb{D}_0^\sharp \models \varphi$. Next, assume that φ is an instance of D2–D6. By the above, we know that $\mathbb{D}_0^\sharp \models \varphi$, as $\mathbb{D}_0^\sharp \models$ D2–D6. Next, assume that φ is an instance of D1. This implies that φ is of the form $D(\psi) \rightarrow \psi$ for some ψ. Then, as $|\psi| \leq l$, we find that $\psi \in \mathbb{D}_0^\sharp$ implies that $\mathbb{D}_0^\sharp \models \psi$ by the induction hypothesis. Thus, if $\mathbb{D}_0^\sharp \models D(\psi)$, then $\mathbb{D}_0^\sharp \models \psi$. As a result, $\mathbb{D}_0^\sharp \models (D(\psi) \rightarrow \psi)$, and, therefore, $\mathbb{D}_0^\sharp \models \varphi$. Finally, assume that φ is of the form $\bigvee_{i=1}^{n} \varphi_i$ for $n > 1$ with some φ_i being an instance of D1–D6. As φ_i is an instance of D1–D6, we know that $\varphi_i \in \mathbb{D}_0^\sharp$, and, as $|\varphi_i| \leq l$, we know by the induction hypothesis that $\varphi_i \in \mathbb{D}_0^\sharp$ implies that $\mathbb{D}_0^\sharp \models \varphi_i$. Thus, we find that $\mathbb{D}_0^\sharp \models \varphi_i$, and, therefore, that $\mathbb{D}_0^\sharp \models \bigvee_{i=1}^{n} \varphi_i$. Thus, in all four cases, we find that $\varphi \in \mathbb{D}_0^\sharp$ implies that $\mathbb{D}_0^\sharp \models \varphi$. Thus, $\mathbb{D}_0^\sharp \models$ D1.

As $\mathbb{D}_0^\sharp \models$ D1–D7, we find that $\mathbb{D}_0^\sharp$ is an ω-model of $T_0^\sharp$. Next, we must show that it is the minimal such ω-model. Assume that $\mathbb{D}$ is an ω-model of $T_0^\sharp$ and that $\varphi \in \mathbb{D}_0^\sharp$. First, assume that $\varphi \in \mathbb{D}_0^\flat$. Then, as $\mathbb{D} \models T_0^\flat$, we find that $\varphi \in \mathbb{D}$ by Proposition A.1. Next, assume that φ is an instance of D1–D6. Then, as $\mathbb{D} \models$ D7, $\varphi \in \mathbb{D}$. Finally, assume that φ is of the form $\bigvee_{i=1}^{n} \varphi_i$, where φ_i is an instance of D1–D6 for some φ_i. Then, as shown above, $\varphi_i \in \mathbb{D}$, and, as $\mathbb{D} \models$ D4, we find that $\bigvee_{i-1}^{i} \varphi_i \in \mathbb{D}$. Thus, for all φ, we find that $\varphi \in \mathbb{D}_0^\sharp$ implies that $\varphi \in \mathbb{D}$. As a result, $\mathbb{D}_0^\sharp \subseteq \mathbb{D}$ for any ω-model $\mathbb{D}$ of $T_0^\sharp$. Thus, we establish that $\mathbb{D}_0^\sharp$ is the minimal ω-model of $T_0^\sharp$. □

By constructing the minimal model of ADT for any consistent deductively closed extension of PA$[D]$, we thereby establish the consistency of ADT, given the consistency of PA$[D]$. ADT thus represents a technically acceptable solution to the alethic paradoxes.

Appendix B: Conservativity of ADT

Though Scharp does not comment on the conservativeness debate in relation to ADT, it is interesting to note that ADT is conservative over strong systems like set theory. Thus, the mere addition of ascending and descending truth predicates to a language does not increase the proof-theoretic strength of such theories. It is presently unclear to us whether this result holds for first-order Peano arithmetic, although the obvious modification of the proof below also works for second-order Peano arithmetic, third-order Peano arithmetic, and so on.

For this proof we will work with the theory ZFC$[D]$, which is simply ZFC with the replacement and separation axiom schemas expanded to include sentences with the descending truth predicate. Furthermore, for the duration of this section, ADT will be axioms D1–D5 + D6$'$ + D7, where D6$'$ is "$D(\varphi)$ if φ is a theorem of ZFC with the replacement and separation schemas for the language of ZFC supplemented with a D predicate." Note that this updated version of D6 simply amounts to replacing PA in the original formulation of the axiom with the now relevant theory ZFC. The idea in the proof below is to focus attention on the restriction of D1–D6$'$ to $L[D]$-sentences in Γ_n, which we define to be Boolean combinations of Π_n or Σ_n-sentences. By this is meant the restriction to the instances of D1–D6$'$ in which the relevant sentences φ, ψ are in the complexity class Γ_n. We abbreviate this class of sentences by $(\text{D1–D6}')_n$. Likewise D7$_n$ is the restriction of D7 to $D(\varphi)$ where φ is from $(\text{D1–D6}')_n$.

The main idea behind these model constructions is to consider the sentences that would be true on all possible extensions of the descending truth predicate. Obviously, the descending truth predicate in the minimal model of ZFC$[D]$ + ADT must include at least these truths. We can again wonder how much bigger the extension of the descending truth predicate must be, and again it turns out that it need not be much bigger, also including instances of the axioms D1–D6$'$ and finite disjunctions with an instance of D1–D6$'$ as one of its disjuncts. Note that the model construction carried out in Proposition B.2 is a natural modification of the construction found in Proposition A.3, altered to function in the setting of ZFC$[D]$ instead of PA$[D]$.

Proposition B.1 *Suppose that M is a model of* ZFC*, and let $P = (P(\omega))^M$ be the powerset of ω relative to M. Define*

$$\mathbb{D}_n^\flat = \big\{\varphi \in \Gamma_n^0 : \forall\, \mathbb{D} \in P\ (M, \mathbb{D}) \models \varphi\big\}. \tag{B.1}$$

Then $(M, \mathbb{D}_n^\flat)$ is a model of ZFC$[D]$ *plus* $(\text{D1–D6}')_n$.

Proof First note that $\mathbb{D}_n^\flat$ is $\Pi_{n+1}^{0,M}$-definable and so is a member of P by separation in M. Hence, we have that for $\varphi \in \Gamma_n$,

$$\varphi \in \mathbb{D}_n^\flat \Longrightarrow (M, \mathbb{D}_n^\flat) \models \varphi. \tag{B.2}$$

From this it follows immediately that D2 holds when restricted to Γ_n-sentences. Further, D3, D4, D5, and D6$'$ hold by definition of $\mathbb{D}_n^\flat$, again when restricted to Γ_n-sentences in the signature $L[D]$. Finally, D1 holds by (B.2). □

Proposition B.2 *Suppose that M is a model of* ZFC*, and let $P = (P(\omega))^M$ be the powerset of ω relative to M. Define $\mathbb{D}_n^\flat$ as in (B.1), and define $\mathbb{D}_n^\sharp$ by*

$$\begin{aligned}\mathbb{D}_n^\sharp = \mathbb{D}_n^\flat &\cup \big\{\varphi \in (\text{D1–D6}')_n\big\} \\ &\cup \Big\{\ulcorner \bigvee_{i=1}^{m} \varphi_i \urcorner \in \Gamma_n : \varphi_i \in (\text{D1–D6}')_n \\ &\qquad \textit{for some } m > 1 \textit{ and } 1 \le i \le m\Big\}.\end{aligned} \tag{B.3}$$

Then $(M, \mathbb{D}_n^\sharp)$ is a model of ZFC$[D]$ *plus* $(\text{D1–D6}')_n$ *and* D7$_n$.

Proof By the second component in the definition of $\mathbb{D}_n^\sharp$, we have that $(M, \mathbb{D}_n^\sharp)$ models D7$_n$. Prior to verifying D2, note that if $\varphi \in \Gamma_n$ and $\varphi \in \mathbb{D}_n^\sharp$, then φ is true on $(M, \mathbb{D}_n^\flat)$. From this it follows that $(M, \mathbb{D}_n^\sharp)$ models $(\text{D2})_n$. The arguments for D3$_n$

and $\mathrm{D4}_n$ are by cases much like the analogous parts of Proposition A.3, whereas those for $\mathrm{D5}_n$ and $\mathrm{D6}'_n$ go through $\mathbb{D}^{\flat}_n$.

Finally, we have the argument for $\mathrm{D1}_n$, which as in Proposition A.3 is by induction on the length of $\varphi \in \Gamma_n$. Where $|\varphi|$ indicates the length of formula φ, we will prove by induction on the length of formulas l that $|\varphi| \leq l \to [\varphi \in \mathbb{D}^{\sharp}_n \to \mathbb{D}^{\sharp}_n \models \varphi]$. As there are no formulas of length 0, clearly this holds for $l = 0$.

Assume that $\varphi \in \mathbb{D}^{\sharp}_n$, that $|\varphi| = l + 1$, and that the induction hypothesis holds for formulas of length $n \leq l$. First, assume that $\varphi \in \mathbb{D}^{\flat}_n$. But by definition of $\mathbb{D}^{\flat}_n$ in (B.1) and the fact that $\mathbb{D}^{\sharp}_n$ is likewise in P, this implies that $\mathbb{D}^{\sharp}_n \models \varphi$. Next, assume that φ is an instance of $(\mathrm{D2}\text{–}\mathrm{D6})_n$. By the argument in the above paragraphs, we know that $\mathbb{D}^{\sharp}_n \models \varphi$, as $\mathbb{D}^{\sharp}_n \models (\mathrm{D2}\text{–}\mathrm{D6})_n$. Next, assume that φ is an instance of D1. This implies that φ is of the form $D(\psi) \to \psi$ for some $\psi \in \Gamma_n$. Then, as $|\psi| \leq l$, we find that $\psi \in \mathbb{D}^{\sharp}_n$ implies that $\mathbb{D}^{\sharp}_n \models \psi$ by the induction hypothesis. Thus, if $\mathbb{D}^{\sharp}_n \models D(\psi)$, then $\mathbb{D}^{\sharp}_n \models \psi$. As a result, $\mathbb{D}^{\sharp}_n \models (D(\psi) \to \psi)$, and, therefore, $\mathbb{D}^{\sharp}_n \models \varphi$. Finally, assume that $\varphi \in \Gamma_n$ is of the form $\bigvee_{i=1}^{m} \varphi_i$ with some φ_i being an instance of $(\mathrm{D1}\text{–}\mathrm{D6})_n$. As φ_i is an instance of $(\mathrm{D1}\text{–}\mathrm{D6})_n$, we know that $\varphi_i \in \mathbb{D}^{\sharp}_n$, and, as $|\varphi| \leq l$, we know by the induction hypothesis that $\varphi_i \in \mathbb{D}^{\sharp}_n$ implies that $\mathbb{D}^{\sharp}_n \models \varphi_i$. Thus, we find that $\mathbb{D}^{\sharp}_n \models \varphi_i$, and, therefore, that $\mathbb{D}^{\sharp}_n \models \bigvee_{i=1}^{m} \varphi_i$. Thus, in all four cases, we find that $\varphi \in \mathbb{D}^{\sharp}_n$ implies that $\mathbb{D}^{\sharp}_n \models \varphi$. Thus, $\mathbb{D}^{\sharp}_n \models \mathrm{D1}_n$. □

Corollary B.1 *Suppose that φ is a sentence in the signature of* ZFC. *Suppose that* $\mathrm{ZFC}[D] + \mathrm{ADT} \vdash \varphi$. *Then* $\mathrm{ZFC} \vdash \varphi$.

Proof Since $\mathrm{ZFC}[D] + \mathrm{ADT} \vdash \varphi$, by compactness φ is deducible from $\mathrm{ZFC}[D]$ plus $(\mathrm{D1}\text{–}\mathrm{D6}')_n \wedge \mathrm{D7}_n$ for some n. Suppose that $\mathrm{ZFC} \nvdash \varphi$. Then by completeness there is a model M of $\mathrm{ZFC} + \neg\varphi$. But by the previous proposition we can expand M into a model $(M, \mathbb{D}^{\sharp}_n)$ of $\mathrm{ZFC}[D]$ plus $(\mathrm{D1}\text{–}\mathrm{D6}')_n \wedge \mathrm{D7}_n$, which contradicts that this theory proves φ while the model satisfies $\neg\varphi$. □

Notes

1. For examples of inconsistency approaches which retain the concept of truth, see Eklund [2], Ludwig [7], and Patterson [9].

2. See Scharp [13] for an extended development of his formal theory. See also Scharp [14] for another development of this theory.

3. For the literature on conservativity, see Field [3] and Shapiro [16]. For the literature on inconsistency approaches, see [2], [7], Priest [11], and Yablo [17].

4. See Priest [10] for an example of an inconsistency view with this virtue.

5. ADT formally includes, in addition to the following seven axioms, axioms for ascending-truth and safety. Due to the definibility of these concepts in terms of descending truth, their inclusion or exclusion does not alter the strength of the theory (see [13, p. 154]). Additionally, Scharp includes an axiom E1 that states that if $s = t$, then $\mathrm{ADT} \models D(s) \leftrightarrow D(t)$. It is not entirely clear where $s = t$ is being evaluated, and

therefore what this axiom entails. For our purposes here, we will thus focus on ADT as being the theory of D1–D7 alone.

6. Scharp states D6 as "$D(\ulcorner\varphi\urcorner)$ if φ is a theorem of PA." It is unclear whether he intends the induction schema to be in the language of PA or the language of PA extended with a descending truth predicate D. For our purposes here, we will consider D6 as including the induction schema in the expanded language. See Section 5 for more on the induction schema in ADT.

7. For a proof of the consistency of ADT using modal logic, see Sharp [13, pp. 157–69] or [15].

8. These arguments originate in the discussion of truth's expressive role in Quine [12]; see [12, p. 12] especially.

9. See Mostowski [8, p. 107] and Shapiro [16, p. 505] for particularly influential defenses of such claims.

10. For a standard treatment of global reflection principles, see Halbach [5, p. 90].

11. See [5, pp. 90–93] for an example of the standard proof of this result.

12. Note that Reflection-A is equivalent to what Scharp calls (Proof-A). See [13, p. 205] for his presentation of the principle. Scharp is not explicit on where the claim that φ is proven is to be evaluated. For our purposes here, we will consider the evaluation to be made in ADT itself.

13. While it is unclear whether ADT can prove Reflection-A, we note that the reflection principle will be true on any ω-model of ADT. See [13, p. 205] for Scharp on the status of Reflection-A.

14. See [5, pp. 90–93] for a standard reference to the formal proof.

15. Note that one trivially could not formulate a reflection principle with the safety predicate, as some unsafe sentences—including the descending liar—are provable.

16. See [8, p. 107] and [16, p. 505].

17. As noted above, it is unclear whether axiom D6 requires that ADT entail all instances of mathematical induction in the language of ADT, or merely in the restricted language of the base theory.

18. We can see this by noting that Scharp intends his concept of safety—as explained above, defined for all φ as $A(\varphi) \rightarrow D(\varphi)$, which is equivalent to the claim that the ascending and descending truth values of ϕ coincide—to serve as a replacement for Kripke's notion of groundedness. See [13, p. 170] for the connection between safety and groundedness, and see Kripke [6] for more on the traditional notion of groundedness.

19. Let the premises $\varphi_1, \ldots, \varphi_n$ all be descending true, let $\varphi_1, \ldots, \varphi_n$ entail ψ by some rule of inference, and let formulas $\varphi_1, \ldots, \varphi_n$ *and* ψ be either descending true or not ascending true. Then $\varphi_1, \ldots, \varphi_n$ entails ψ which entails $A(\psi)$ which entails $D(\psi)$. Thus, the

rules of inference preserve descending and ascending truth as applied to sentences all of which are either descending true or not ascending true. As all sentences in the language of PA have this property according to ADT, we thereby find that all rules of inference as applied to statements in PA are ascending and descending truth preserving.

20. See [15] and [13, pp. 157–69] for a nonconstructive proof of the consistency of ADT. The following proof, alternatively, is more constructive.

References

[1] Dean, W., "Arithmetical reflection and the provability of soundness," *Philosophia Mathematica (3)*, vol. 23 (2015), pp. 31–64. Zbl 06691545. MR 3335259. DOI 10.1093/philmat/nku026. 443, 444

[2] Eklund, M., "Inconsistent languages," *Philosophy and Phenomenological Research*, vol. 64 (2002), pp. 251–75. 451

[3] Field, H., "Deflating the conservativeness argument," *Journal of Philosophy*, vol. 96 (1999), pp. 533–40. MR 1718770. 451

[4] Field, H., *Saving Truth from Paradox*, Oxford University Press, Oxford, 2008. Zbl 1225.03006. MR 2723032. 445

[5] Halbach, V., *Axiomatic Theories of Truth*, Cambridge University Press, Cambridge, 2011. Zbl 1223.03001. MR 2778692. 452

[6] Kripke, S. A., "Outline of a theory of truth," *Journal of Philosophy*, vol. 72 (1975), pp. 690–716. Zbl 0952.03513. DOI 10.2307/2024634. 452

[7] Ludwig, K., "What is the role of a truth theory in a meaning theory?" pp. 142–63 in *Meaning and Truth: Investigations in Philosophical Semantics*, edited by J. K. Campbell, M. O'Rourke, and D. Shier, Seven Bridges, New York, 2002. 451

[8] Mostowski, A., *Sentences Undecidable in Formalized Arithmetic: An Exposition of the Theory of Kurt Gödel*, North-Holland, Amsterdam, 1952. Zbl 0047.00903. MR 0048366. 452

[9] Patterson, D., "Inconsistency theories of semantic paradox," *Philosophy and Phenomenological Research*, vol. 79 (2009), pp. 387–422. 451

[10] Priest, G., "The logic of paradox," *Journal of Philosophical Logic*, vol. 8 (1979), pp. 219–41. Zbl 0402.03012. MR 0535177. DOI 10.1007/BF00258428. 451

[11] Priest, G., *In Contradiction: A Study of the Transconsistent*, vol. 39 of *Nijhoff International Philosophy Series*, Martinus Nijhoff, Dordrecht, 1987. Zbl 0682.03002. MR 1014684. 451

[12] Quine, W. V., *Philosophy of Logic*, 2nd ed., Harvard University Press, Cambridge, Mass., 1986. MR 0844769. 452

[13] Scharp, K., *Replacing Truth*, Oxford University Press, Oxford, 2013. 438, 439, 440, 441, 442, 443, 444, 445, 451, 452, 453

[14] Scharp, K., "Truth, the liar, and relativism," *Philosophical Review*, vol. 122 (2013), pp. 427–510. 451

[15] Scharp, K., "Xeno semantics for ascending and descending truth," pp. 149–67 in *Foundational Adventures: Essays in Honor of Harvey M. Friedman*, edited by N. Tennant, vol. 22 of *Tributes*, College Publications, London, 2014. Zbl 1358.03015. MR 3241958. 452, 453

[16] Shapiro, S., "Proof and truth: Through thick and thin," *Journal of Philosophy*, vol. 95 (1998), pp. 493–521. MR 1651807. 440, 451, 452

[17] Yablo, S., "Definitions, consistent and inconsistent," *Philosophical Studies*, vol. 72 (1993), pp. 147–75. MR 1259703. 451

Acknowledgments

I would like to thank Henri Galinon, Martin Fischer, Rohan French, Sean Walsh, and an anonymous referee for their many helpful comments on earlier versions of this work.

Department of Logic and Philosophy of Science
University of California at Irvine
Irvine, California
USA
schatzj@uci.edu