Intersubstitutivity Principles and the Generalization Function of Truth

Anil Gupta
University of Pittsburgh

Shawn Standefer
University of Melbourne

Abstract We offer a defense of one aspect of Paul Horwich's response to the Liar paradox—more specifically, of his move to preserve classical logic. Horwich's response requires that the full intersubstitutivity of ' 'A' is true' and A be abandoned. It is thus open to the objection, due to Hartry Field, that it undermines the generalization function of truth. We defend Horwich's move by isolating the grade of intersubstitutivity required by the generalization function and by providing a new reading of the biconditionals of the form " 'A' is true iff A."

Keywords Truth · Paradox · Minimalism · Deflationism · Revision theory · Paul Horwich · Hartry field

1. Introduction

In §10 of his influential book *Truth*, Paul Horwich considers how his Minimal Theory (MT) should be modified in face of the Liar paradox. Horwich's initial formulation of MT has it consisting of the T-biconditionals for all sentences *A*, where *the T-biconditional for A* is the sentence

'*A*' is true iff *A*.[1]

---

[1]MT is actually formulated as a theory of propositional truth. For the issues addressed in this paper, however, it is simpler to work with sentential truth. Horwich treats sentential truth and propositional truth in a parallel way (see chapter 7 of *Truth*), and we will feel free to transpose his remarks concerning propositional truth into those concerning sentential truth.

It is not essential to our argument below that truth be treated as a predicate of sentences. The argument could easily be recast into a propositional key, but it would become longer without becoming more illuminating.

We assume the usual conventions necessary for treating truth as a predicate of sentences.

Horwich notes that under this formulation, a contradiction can be derived from MT if there is a Liar sentence in the language. If we set

l = '*l* is not true',

thus making '*l* is not true' a Liar sentence, then, on Horwich's initial formulation, MT contains the biconditional

'*l* is not true' is true iff *l* is not true,

which is inconsistent with the truth that *l* = '*l* is not true'.

Horwich considers four possible ways of responding to the problem: (1) deny classical logic; (2) deny the applicability of truth to sentences like the Liar; (3) deny that Liar sentences are well formed; and (4) exclude some T-biconditionals from MT. He rejects the first three ways and opts for the fourth. Horwich does not provide a precise specification of MT, however. He does not spell out which T-biconditionals are to belong to MT and which are to be excluded. Horwich imposes only the following general constraints on any future specification of MT:

> (a) that the minimal theory not engender 'liar-type' contradictions; (b) that the set of excluded instances be as small as possible; and . . . (c) that there be a constructive specification of the excluded instances that is as simple as possible. (*Truth*, p. 42)

In this paper we offer a partial defense of Horwich's response to the paradoxes. More particularly, we defend Horwich's rejection of option (1), the option to deny classical logic. This move leads Horwich to exclude some T-biconditionals from MT and opens him up to an important objection—an objection against which we shall offer a defense. The objection is that MT does not support the full intersubstitutivity of *A* and ' '*A*' is true' and, consequently, cannot underwrite the generalization function of truth. We explain this objection, which is due to Hartry Field, in section II, and we sketch a way of resisting it in section III. In section IV, we offer a formulation of MT that meets the objection. This formulation satisfies conditions (a)-(c) above, but requires a modification in Horwich's response to the paradoxes.[2]

2. Full Intersubstitutivity

Field has claimed that

> in order for the notion of truth to serve its purposes, we *need* what I've been calling the *Intersubstitutivity Principle*.[3]

Field's intersubstitutivity principle is this:

> **Full/Field Intersubstitutivity (FI)**: Sentences *A* and ' '*A*' is true' are intersubstitutable in all extensional contexts. More precisely, if *B* is an arbitrary sentence and B^* results from *B* by replacing one or more extensional occurrences of the form ' '*A*' is true' in *B* by *A* then *B* and B^* are interderivable (i.e., *B* may be derived from B^*, and B^* from *B*).[4]

The purposes of the notion of truth that Field has in mind were highlighted by W. V. Quine, who

---

[2]Because of the required modification, our defense of Horwich is only partial; we are unable to defend all aspects of Horwich's response.

We should note that our interest in mounting a defense of Horwich does not issue from a desire to support deflationism, which animates Horwich's book. Our interest issues from a desire to support the idea that classical logic is perfectly consistent with the generalization function of truth. Even if we succeed, as we hope, in providing deflationists with a better formulation of MT, they will not thereby be in a better position to answer the main objections to deflationism. For example, the objections given in Anil Gupta's "Critique of Deflationism" are entirely independent of the proper formulation of MT in face of the Liar.

[3]Field, *Saving Truth from Paradox*, p. 210; italics added to 'need'.

[4]In rough terms, a sentential context ' . . . *A* . . .' counts as extensional iff, necessarily, if $A \equiv B$ then (. . . *A* . . . ≡ . . . *B* . . .); here '≡' expresses material equivalence. A similar scheme explains the extensionality of predicate and name positions.

For the purposes of the discussion in this section, we can assume that the language containing the truth-predicate is a first-order language, equipped with quotation names of sentences. In such a language, all contexts are extensional except those occurring within quotation names.

explained them as follows in a famous passage in his *Philosophy of Logic*:

> We may affirm the single sentence by just uttering it, unaided by quotation or by the truth predicate; but if we want to affirm some infinite lot of sentences that we can demarcate only by talking about the sentences, then the truth predicate has its use. We need it to restore the effect of objective reference when for the sake of some generalization we have resorted to semantic ascent.[5]

The point Quine makes about infinite lots of sentences holds also for finite ones. Suppose a friend, Bill, makes a long speech consisting of sentences A_1, . . . , and A_n, and you want to affirm this finite lot of sentences. Assuming that you are in a conversational context in which it is common knowledge what Bill has said, you do not need to repeat Bill's long speech. You can achieve the desired effect more efficiently by "resorting to semantic ascent." You can affirm

(1) Everything Bill says is true.

The truth-predicate enables you to achieve the effect of affirming the long-winded (2),

(2) A_1 & . . . & A_n,

through an affirmation of the compact (1). Now, given common knowledge of what Bill said, affirmation of (1) is equivalent (in some sense of 'equivalent') to an affirmation of

'A_1' is true & . . . & 'A_n' is true.

So, for the notion of truth to serve its function here, an affirmation of ' 'A_i' is true' needs to be equivalent to an affirmation of A_i ($1 \leq i \leq n$).

The exact nature of "equivalent affirmation" in play in the above argument is important

[5] P. 12; see also Quine's *Pursuit of Truth*, §33.

for some purposes—for example, for assessing the correctness of deflationism.[6] For present purposes, however, we can work with a weak notion:

An affirmation of B is equivalent to an affirmation of C

iff B and C are *co-affirmable*; that is,
iff (B is affirmable iff C is affirmable); in symbols:
iff ($\models B$ iff $\models C$).

Now it is plain that if truth is to serve the above kind of generalization function, an intersubstitutivity must hold between A and ' 'A' is true': the two sentences must be intersubstitutable in affirmations. That is, the following principle must hold:

Categorical Intersubstitutivity (CI): $\models$ 'A' is true iff $\models A$.[7]

Field points out that (CI) is insufficient to underwrite the full generalization function of truth. Field observes that 'true' serves a generalization function not only in categorical contexts but also in embedded contexts—for example, when truth-attributions occur as antecedents of conditionals. Consider a variant of the first example above. Suppose Bill makes, as before, a long speech consisting of $A_1, \ldots,$ and A_n. You, however, do not wish to affirm $A_1, \ldots,$ and A_n. You wish to affirm, instead, a conditional with (A_1 & . . . & A_n) as its antecedent. Say, you wish to affirm

(3) If A_1 & . . . & A_n then Bernie will win.

[6]See "Critique of Deflationism."

[7]According to Quine, the function of the truth-predicate is to enable one to express generalizations over sentence positions (e.g., the position of A in 'if A then A') using nominal quantification (as in, e.g., "for all sentences x, x is true if x is a conditional whose antecedent and consequent are identical"). The truth-predicate can play this expressive role *only if* certain intersubstitutivity principles hold. (CI) is one such principle. We are concerned to discover which further principles are required by this expressive role.

Assuming, as before, that you are in a conversational context in which it is common knowledge what Bill has said, you can achieve your goal by affirming the compact (4) instead of the cumbersome (3).

(4) If everything Bill says is true then Bernie will win.

We have here a perfectly ordinary, legitimate use of the truth-predicate. Principle (CI), however, is insufficient to underwrite it. An affirmation of (4), we can grant, amounts to an affirmation of (5):

(5) If ‘A_1’ is true & . . . & ‘A_n’ is true then Bernie will win.

The difficulty is that (CI) warrants intersubstitutivity of *A* and ‘ ‘*A*’ is true’ only in categorical contexts, when one of these sentences is affirmed, not in hypothetical contexts such as (3) and (5). We need a stronger intersubstitutivity principle. Field points out that (FI) suffices here, though (CI) does not. Field thinks that (FI) not only suffices but also that it is *needed*.

If Field is right that the generalization function requires (FI), then it requires also that classical logic be abandoned. Classically,

(6) $A \equiv A$

is a logical law (here ‘≡’ expresses material equivalence), and (FI) renders this law interderivable with

(7) ‘*A*’ is true $\equiv A$,

which, as we saw above, implies an inconsistency in the presence of a Liar sentence. If we wish to preserve both consistency and the generalization function of truth, it follows that we must abandon classical logic.[8] Indeed, Field embraces precisely this course as the way out of the problem.[9]

3. Uniform Intersubstitutivity

We can begin to resist Field's claim by observing that the co-affirmability of (3) and (5) does not require (FI). A much weaker principle suffices, which we may formulate thus:

> **Uniform Intersubstitutivity (UI)**: Let B be an arbitrary sentence in which all extensional occurrences of the truth-predicate are confined to contexts of the form ' 'A' is true'. Let B^* be the sentence that results when *each* extensional occurrence of the form ' 'A' is true' in B is replaced by A. Then, B and B* are co-affirmable:
>
> $\models B$ iff $\models B^*$.

Full Intersubstitutivity (FI) allows some occurrences of ' 'A_i' is true' ($1 \leq i \leq n$) in (5) to be replaced by A_i while leaving others unchanged. Uniform Intersubstitutivity, on the other hand, requires that if some ' 'A_i' is true' in (5) are replaced by A_i ($1 \leq i \leq n$), then *all* of them should be

---

[8]Horwich's response to the problem is to keep classical logic and to yield on the generalization function of truth. Since Horwich restricts even (CI), truth will not serve the generalization function in full. Horwich thinks this is not a cause for concern, since "the utility of truth as a device of generalization is not substantially impaired" (*Truth*, p. 42, fn. 21). Whether the impairment is substantial or not depends, however, on the details of the minimal theory, which Horwich has not supplied. If the T-biconditionals excluded from MT are few, then the damage may well be minimal; otherwise, the damage could be substantial. There is, in either case, an additional difficulty: MT will not be able to explain how the sentence 'a conjunction is true only if its first conjunct is true' generalizes truths of the form 'if A and B then A'.

[9]Even within a three-valued logic, which Field favors, (FI) implies that truth can serve the generalization function only at the cost of expressive richness. If, for example, exclusion negation is expressible, then the same difficulty arises in the three-valued context as in the classical. If Field is right about the necessity of (FI), then the full generalization function of truth can be obtained only at the expense of some expressive resources. The argument below is aimed at resisting this conclusion. We think that the full generalization function of truth can be had without limiting logical or non-logical resources in any way.

replaced. The extra freedom provided by (FI) does not seem to be needed for the generalization function of truth. It is certainly not needed to ensure that the required equivalence obtains between (3) and (5). Furthermore, it is this extra freedom that generates a problem with classical logic. The move from (6) to (7) requires (FI). (UI) allows us to move only to the innocuous,

‘*A*’ is true ≡ ‘*A*’ is true.

This suggests that the generalization function of truth can be preserved without weakening classical logic. Perhaps all that the generalization function requires is (UI), not (FI). We say ‘perhaps’ because the literature on the concept of truth does not provide an exact characterization of the generalization function, and no exact characterization is easily forthcoming. Nonetheless, the following observation is evidence that (UI) is all one needs.

Consider again the example above, in which the conversational context renders (1) and (2) equivalent.

(1) Everything Bill says is true.

(2) A_1 & . . . & A_n.

Suppose we say that the generalization function of truth requires that the effect of an affirmation of any truth-functional compound that contains occurrences of (2) should be achievable using (1). So, for instance, we should be able, for any sentences *B* and *C*, to get the effect of affirming

(8) [(A_1 & . . . & A_n) & *B*] or [not (A_1 & . . . & A_n) & *C*]

by affirming a truth-functional compound containing (1). Now, we claim that (UI) suffices to sustain this sort of generalization function; (FI) is not needed.

Observe that we cannot get the effect of affirming (8) by affirming (9):

(9) [Everything Bill says is true & *B*] or [not everything Bill says is true & *C*].

Our assumptions render (9) equivalent to (10):

(10) [('A_1' is true & . . . & 'A_n' is true) & *B*] or [not ('A_1' is true & . . . & 'A_n' is true) & *C*].

The difficulty is that (UI) does not ensure that (8) and (10) are co-affirmable, since *B* and *C* may contain truth attributions. Furthermore, (UI) is right not to declare (8) and (10) co-affirmable, for there are instances of (8) and (10) that are *not* co-affirmable. We can obtain such an instance if we set $n = 1$ and let A_1 and *B* be the Liar sentence '*l* is not true' and *C* the negation of the Liar sentence. Now, (8) is equivalent to a tautology, but (10) is equivalent to a contradiction. Note that this example shows that the generalization function should not be read as implying the intersubstitutivity of (1) and (2) in all extensional contexts.

Fortunately, there is another way of obtaining the effect of affirming (8). We can affirm (11):

(11) [Everything Bill says is true & '*B*' is true] or
[not everything Bill says is true & '*C*' is true].

Our assumptions render (11) equivalent to (12):

(12) [('A_1' is true & . . . & 'A_n' is true) & '*B*' is true] or
[not ('A_1' is true & . . . & 'A_n' is true) & '*C*' is true].

Principle (UI) ensures that (8) and (12) are co-affirmable. The principle thus allows us to use semantic ascent to get the effect of affirming the long-winded (8) through an affirmation of the compact (11).[10] We conclude that there is reason to think that (UI) suffices for the generalization

[10] If (2) occurs embedded within quantifiers, then the above technique will work only in some cases, not in all. The technique will not work if one of these quantifiers, say 'for all *x*', binds an occurrence of '*x*' within '*x* is true'. For such cases, semantic ascent, instituted as outlined above, requires the satisfaction predicate. An analog of (UI) formulated for the satisfaction predicate gives us the needed intersubstitutivities. We do not know what is possible here merely with the truth predicate.

function of truth.[11]

But does (UI) provide a tenable middle ground between (FI) and (CI) over which classical logic can remain in force? (UI) requires that we recognize a substantial connection between every sentence *A* and its truth-attribution, ' '*A*' is true'. Without this connection, we cannot guarantee the co-affirmability of (3) and (5), for there is no restriction on what the A_i's in (3) may be.[12] On the other hand, we cannot recognize this connection via the *material T-biconditionals*

'*A*' is true ≡ *A*,

for these yield contradictions in the presence of classical logic. What sort of connection might there be between ' '*A*' is true' and *A* that is strong enough to sustain (UI), but not so strong that it brings (FI) in its wake? And how might this connection be expressed? That is, how might we formulate the theory of truth so that it institutes a connection between ' '*A*' is true' and *A* that sustains (UI) but not (FI)?

The key to the answer lies in a remark of Alfred Tarski's.

4. The Proper Formulation of the Minimal Theory

Tarski said of the T-biconditional that it "may be considered a *partial definition* of truth, which

---

[11] A couple of points of clarification concerning (UI):

(i) Our principal conclusions do not rest on the claim about the sufficiency of (UI). So long as the needed intersubstitutivity principles are consistent with classical logic, our principal conclusions hold.

(ii) We do not take (UI) to be a full theory of truth, one that suffices to explain our uses of the concept of truth. There are principles governing truth that go beyond (UI), for example, the T-biconditionals under the reading proposed below.

Thanks to an anonymous referee for pushing us to clarify the role of (UI) in our argument.

[12] The generalization function remains intact even if some of the A_i's are paradoxical. You can gain the effect of affirming (3) by affirming (4), even when some of the A_i's are paradoxical.

explains wherein the truth of . . . one individual sentence consists."[13] If this is right, as we think it is, then the T-biconditional is often a *circular* partial definition of truth. The T-biconditional of, e.g., 'everything Bill says is true' explains the conditions under which this sentence is true in terms that use the notion of truth itself—the right-hand side of following T-biconditional contains the truth-predicate:

(13) 'Everything Bill says is true' is true iff everything Bill says is true.

Now, revision theory provides tools for making sense of circular definitions, and one lesson of this theory is that, with circular definitions, a sharp distinction must be made between definitional and material readings of 'iff': (13) read definitionally does not imply (13) read materially, and the latter reading does not imply the former.

Classical revision theory, though it highlights the distinction between the two readings of 'iff', does not provide resources for expressing anything like the definitional reading in a language containing truth (and, more generally, circular concepts). In recent work, we have shown how revision theory can be extended to yield languages that contain two different biconditionals: the material biconditional (≡) and, what we call, the *step biconditional* (↔).[14] This latter connective provides, we think, the resource needed for solving the problem before us—the problem of connecting ' '*A*' is true' with *A* in a way that is strong enough to sustain (UI) but not so strong that it commits one to (FI). Let *the step T-biconditional for A* be the sentence

'*A*' is true ↔ *A*.

We show that the step T-biconditionals sustain (UI) but not (FI).[15]

[13] Tarski, "Semantic Conception of Truth and the Foundations of Semantics," §4; italics added.

[14] See our "Conditionals in Theories of Truth," where we define the step biconditional using two step-conditionals. For further information about the logic of these conditionals, see Shawn Standefer's "Solovay-Type Theorems for Circular Definitions."

[15] In the exposition below, we aim to impart intuitive understanding, without bringing into play the technical details of revision theory. For these details, see Gupta and Nuel Belnap's *Revision Theory of Truth* and our "Conditionals in Theories of Truth."

We begin with an intuitive sketch of the revision-theoretic interpretation of the step biconditional. In revision theory, a circular definition is interpreted via a revision process—a process consisting of successive stages in which the interpretation of the defined term can shift from stage to stage.[16] Suppose that the defined term is a one-place predicate *G*. Then, the rule that determines the shifting interpretations is this: the interpretation of *G* at a successor stage $\alpha + 1$ consists of those objects that satisfy the definiens when *G* is assigned the stage α interpretation.[17] Thus, the circular definition serves as a rule for revising *G*'s interpretation. A sentence *B* is *affirmable* ($\models B$) iff it is almost everywhere true in the revision process; that is, iff at almost all revision stages, *B* is true.[18] It is a fact that

$\models B$ iff, for almost all stages α, *B* is true at α;
iff, for almost all stages α, *B* is true at the successor stage $\alpha + 1$; and
iff, for almost all stages α, *B* is true at the successor stage $\alpha + 2$.

Now, the step biconditional is governed by the following rule:

$(A \leftrightarrow B)$ is true at a stage $\alpha + 1$ iff [*A* is true at stage $\alpha + 1$ iff *B* is true at stage α].

The point to note is that the step biconditional $(A \leftrightarrow B)$ expresses a *cross-stage* connection. Its affirmability requires that, at almost all revision stages α, *B*'s truth-value at α should coincide with *A*'s truth-value at the *next* stage, $\alpha + 1$. The material biconditional $(A \equiv B)$, in contrast, expresses a *same-stage* connection. Its affirmability requires that, at almost all stages α, the truth-value of *B* at stage α coincides with the truth-value of *A* at the *same* stage α. In the special case in which the truth-values of *A* and *B* do not fluctuate in the revision process (or do not fluctuate much), the two biconditionals are co-affirmable:

---

[16]Technically, the revision process is the collection of sequences one obtains through repeated revisions of arbitrary revision hypotheses. A stage in the process is simply a stage in one of these sequences.

[17]There are rules for assigning *G* an interpretation at limit stages; see Gupta and Belnap, *Revision Theory of Truth*.

[18]We are using the rough notion "almost everywhere true" in place of "valid in $\mathbf{S}^{\#}$."

$\models (A \leftrightarrow B)$ iff $\models (A \equiv B)$.

In general, though, the above equivalence fails.

Let us turn now to the revision process for truth. The interpretation of the truth-predicate at a successor stage $\alpha + 1$ consists of the sentences assessed as true when the truth-predicate receives the stage α interpretation. The interpretation of the truth-predicate is classical at each stage, and consequently, classical logical laws are assessed as true at each stage; these laws are all affirmable.[19] Now, the revision rule for truth ensures that at all stages α,

‘ ‘*A*’ is true’ is true at stage $\alpha + 1$ iff *A* is true at stage α.

Hence, we have

‘ ‘*A*’ is true $\leftrightarrow$ *A*’ is true at stage $\alpha + 1$, and
‘ ‘ ‘*A*’ is true $\leftrightarrow$ *A*’ is true’ is true at stage $\alpha + 2$.

It follows that

$\models$ ‘*A*’ is true $\leftrightarrow$ *A*, and
$\models$ ‘ ‘*A*’ is true $\leftrightarrow$ *A*’ is true.

That is, the step T-biconditionals are all affirmable, as also are the attributions of truth to them.

It is a feature of the revision process for truth that the truth-values of ordinary, non-paradoxical sentences *B* do not fluctuate very much, and the same holds of their truth-attributions, ‘ ‘*B*’ is true’. For such sentences *B*, we have

$\models$ ‘*B*’ is true $\leftrightarrow$ *B* iff $\models$ ‘*B*’ is true $\equiv$ *B*.

[19] This is so because we have chosen to work within a classical framework. Revision processes for truth can be constructed for non-classical languages also.

So, the material T-biconditionals for these sentences are affirmable also.[20] The distinction between the two biconditionals, material and step, is of little significance over ordinary, non-paradoxical sentences.

The truth-value of a paradoxical sentence fluctuates through the revision process. For example, if the Liar sentence '*l* is not true' is true at one stage then it is false at the next, and true again at the following stage; and the pattern repeats *ad infinitum*. Consequently, at almost every stage α, '*l* is not true' and the attribution of truth to it possess opposite truth-values. The negation of the material T-biconditional of the Liar sentence is thus affirmable, even though the step T-biconditional of the Liar sentence, like those of every sentence, is affirmable:

$\models$ ~['*l* is not true' is true ≡ *l* is not true], and
$\models$ '*l* is not true' is true ↔ *l* is not true.

It follows that Full Intersubstitutivity (FI) fails, since it requires the affirmability of the material T-biconditional of the Liar.

Uniform Intersubstitutivity (UI) holds, however. We show this for a simplified version (3) and (5), but the argument is easily generalized. Let *B* be a sentence without any occurrences of the truth-predicate (e.g., 'Bernie will win') and let '⊃' be the material conditional. Then, we show that

(14) $\models$ (*C* & *D*) ⊃ *B* iff $\models$ ('*C*' is true & '*D*' is true) ⊃ *B*.

Suppose *B* is true at a successor stage $\alpha + 1$. Then *B* is true at all successor stages $\alpha + 1$ (since *B* has no occurrences of the truth-predicate, and only the interpretation of the truth-predicate changes from stage to stage). Hence, at all stages $\alpha + 1$, both

(*C* & *D*) ⊃ *B*,

[20]The attributions of truth to these material T-biconditionals are also affirmable.

(‘*C*’ is true & ‘*D*’ is true) ⊃ *B*

are true. So, (14) holds. Suppose, on the other hand, that *B* is false at a successor stage $\alpha + 1$. Then, for the same reason as before, *B* is false at all successor stages $\alpha + 1$. Now the following equivalences hold at almost all stages α,

‘(*C* & *D*) ⊃ *B*’ is true at $\alpha + 1$

iff one of *C* or *D* is false at $\alpha + 1$,

iff one of ‘ ‘*C*’ is true’ or ‘ ‘*D*’ is true’ is false at $\alpha + 2$,

iff ‘(‘*C*’ is true & ‘*D*’ is true) ⊃ *B*’ is true at $\alpha + 2$.

Consequently,

‘(*C* & *D*) ⊃ *B*’ is true at almost all revision stages

iff ‘(‘*C*’ is true & ‘*D*’ is true) ⊃ *B*’ is true at almost all revision stages.

Hence, (14) holds in this case as well.[21]

So, classical logic and (UI) can both be sustained if we view the link between ‘ ‘*A*’ is true’ and *A* as definitional. The definitional links for circular concepts cannot be expressed through the material biconditional; they require us to bring in play the step biconditional.

One more observation: there cannot be a constructive specification of the material T-biconditionals that are affirmable, at least not if ‘constructive’ is understood as imposing a meaningful constraint. Let *Q* be a false statement, say, of arithmetic or set theory. Then, the conjunction of *Q* with the Liar sentence,

Q & (*l* is not true),

[21] A relevant theorem: If *B* and *B** meet the conditions laid down in (UI), then *B* is deducible from the step T-biconditionals in the calculus $\mathbf{C}_0^+$ iff *B** is also so deducible. See our “Conditionals in Theories of Truth” for the details of $\mathbf{C}_0^+$.

is false (intuitively, as well as at almost all revision stages), and so also is the attribution of truth to it. We have, therefore,

$\models$ '*Q* & (*l* is not true)' is true ≡ *Q* & (*l* is not true).

(Indeed, this material T-biconditional can be derived from the step T-biconditional for '*Q* & (*l* is not true)'.) On the other hand, if *Q* is true then the material T-biconditional for '*Q* & (*l* is not true)' is equivalent to that of the Liar sentence, and is therefore not affirmable. It follows that if there were a constructive specification of the affirmable material T-biconditionals, then there would be a constructive specification of the arithmetical truths, of set-theoretical truths—and, indeed, of all truths.

In summary, then, we have argued that the generalization function of truth does not require Full Intersubstitutivity. We formulated a weaker condition, Uniform Intersubstitutivity, whose truth is sustained by the step T-biconditionals and which, we suggested, suffices for the generalization function of truth. There is, thus, a formulation of the Minimal Theory, MT, that preserves classical logic and that implies the needed intersubstitutivities. Horwich's unrestricted MT implies the needed intersubstitutivities but leads to inconsistencies in the presence of classical logic. Horwich's restricted MT preserves classical logic but does not imply the needed intersubstitutivities. We have shown how to formulate MT so that it both preserves classical logic and implies the needed intersubstitutivities. Horwich's move to preserve classical logic is, thus, perfectly viable. It requires, though, that the Minimal Theory consist not of a selection of the material T-biconditionals, but of *all* the step T-biconditionals.[22] This way we engender no contradictions; we exclude T-biconditionals for no sentences; and we gain not only a constructive but the simplest possible specification of the axioms of the Minimal Theory. We meet, that is, Horwich's conditions (a)-(c) quoted in section I.

Acknowledgements We are grateful to the editors, Joseph Ulatowski and Cory Wright, and particularly to an anonymous referee for their helpful comments on our paper. Standefer's acknowledgement: This research was supported by the Australian Research Council, Discovery Grant, DP150103801.

[22]Note that classical revision theory does not provide the resources—more specifically, a suitable sentential connective—for a satisfactory formulation of the Minimal Theory.

References

Gupta, Anil. "A Critique of Deflationism." Reprinted with a postscript in his *Truth, Meaning, Experience*, pp. 9-52. New York: Oxford University Press, 2011. Originally published in 1993.

Gupta, Anil and Belnap, Nuel. *The Revision Theory of Truth*. Cambridge MA: MIT Press, 1993.

Gupta, Anil and Standefer, Shawn. "Conditionals in Theories of Truth." *Journal of Philosophical Logic* (2017). doi:10.1007/s10992-015-9393-3.

Horwich, Paul. *Truth*. Second edition. Oxford: Clarendon Press, 1998. The first edition appeared in 1990.

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

Quine, W. V. *Philosophy of Logic*. Second edition. Cambridge MA: Harvard University Press, 1986. The first edition appeared in 1970.

Quine, W. V. *Pursuit of Truth*. Revised edition. Cambridge MA: Harvard University Press, 1992. The original edition appeared in 1990.

Standefer, Shawn. "Solovay-Type Theorems for Circular Definitions." *Review of Symbolic Logic* 8 (2015), 467-487.

Tarski, Alfred. "The Semantic Conception of Truth and the Foundations of Semantics." Reprinted in Lynch, Michael P., *The Nature of Truth: Classical and Contemporary Perspectives*, pp. 331-363. Cambridge MA: MIT Press, 2001. Originally published in 1944.