The Liar Paradox: Tangles and Chains
Author(s): Tyler Burge
Source: *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, Vol. 41, No. 3 (May, 1982), pp. 353-366
Published by: Springer
Stable URL: http://www.jstor.org/stable/4319529
Accessed: 11-04-2017 02:26 UTC

TYLER BURGE

THE LIAR PARADOX: TANGLES AND CHAINS

(Received 20 March, 1981)

In this paper I want to use the theory that I applied in 'Semantical paradox' to the Liar and Grelling paradoxes to handle some subtler versions of the Liar.[1] These new applications will serve as tools for sharpening some of the pragmatic principles and part of the underlying motivation of the theory. They also illustrate the superiority of our theory to previous ones in the hierarchical tradition.

I

The heart of the theory is that 'true' is an *indexical-schematic* predicate. Most of its uses are indexical; some are schematic. A natural language predicate on an occasion of use is *indexical* if its extension on that occasion depends not only on the contextually appropriate conventional meaning of the predicate, but also on the context of its use. A natural language predicate on an occasion of use is *schematic* if it lacks a definite extension on that occasion, but through its use on that occasion provides general constraints on the extension of the same predicate on other occasions of use.

Part of the point of taking 'true' to be indexical may be seen by considering the Liar:

(1) (1) is not true.

We know that if we substitute (1) in a Tarskian truth-schema with an occurrence of 'true' having the same extension as the occurrence in (1), we get a contradiction. This leads us to conclude that such a schema will not apply to (1). This is to say that (1) lacks conditions for being true (where 'true' in this last occurrence is extensionally equivalent with 'true' as it occurs in (1)). Intuitively it follows that whatever may be wrong with (1), (1) is not true. (It may not be false either: I shall not discuss this matter here.) But now we have just asserted '(1) is not true'. In this assertion, '(1)' refers to (1) and

Philosophical Studies 41 (1982) 353–366. 0031-8116/82/0413–0353$01.40

'true' is extensionally equivalent with 'true' as it occurs in the original token of (1) set off above. Thus (assuming negation remains constant) we have just asserted the same sentence, under the same interpretation, that we are counting not true. Seemingly, we have done so reasonably. We are committed to the truth of what we assert. So '(1) is not true' (taken in the original interpretation) is true – which is to say '(1) is true'.

Have we contradicted ourselves? No, we have been entirely reasonable. What has happened is that 'true' has undergone a shift of extension (without changing its meaning) in the course of our argument. The occurrence of 'true' in the original token of (1) has a different, narrower extension than the occurrence of 'true' in '(1) is true' – the sentence that concludes the reasoning. In brief, 'true' is indexical on these occasions of use.

A thorough discussion of the details of this account will not be given here. But the fundamental idea is that in attempting to insert (1) into a truth schema with an occurrence of 'true' univocal (extensionally equivalent) with that in (1) itself, we come to see (as a result of producing a contradiction) that a pragmatic implicature associated with the initial occurrence of (1) does not hold. Call the original occurrence of (1) 'an occurrence of *the evaluating sentence*'. Call a truth schema whose occurrence of 'true' is univocal with 'true' in the occurrence of the evaluating sentence 'the truth schema *associated with* the occurrence of the evaluating sentence'. Then the implicature in its most general form is this:

(I) Sentences (under a contextually determined interpretation) which are referred to or quantified over in the occurrence of the evaluating sentence are to be given truth-conditions – schematically evaluated – by means of the truth schema associated with the occurrence of the evaluating sentence.

This implicature usually holds. But it fails in the case of the original occurrence of (1) because attempting to maintain it leads to contradiction. The implicature is cancelled. The general reason behind all failures of the implicature (including many that are not precipitated by contradictions) is that accepting the implicature in the context would violate a fundamental condition on semantical evaluation: The semantical value (e.g., the truth value) of the expression being evaluated should be fixed independently of the evaluation itself; the evaluation should not depend on its own outcome. Thus, if (1) were evaluated via its associated truth schema, its semantical value could be fixed only by considering whether the truth predicate evaluating it applied or not. If (a), '(a) is true', were evaluated by its associated truth schema, it would be true if it were true – and otherwise not. The evaluation would be unacceptably empty or circular.

When we reject implicature (I), we conclude, "(1) [interpreted as it is in its original occurrence] is not true" – not because (1) at that occurrence has truth conditions which fail to obtain, but because it *lacks truth conditions*. It is no more true than 'is red' is true. In this reasoning, saying that the initial occurrence 'lacks truth conditions' is merely an abbreviated way of saying that it is not appropriately substitutable into its *associated* truth schema.

We have concluded '(1) is not true'. The occurrence of 'true' in this conclusion is still univocal with its occurrence in the preceding reasoning. (Otherwise, we will not be judging the original application of the predicate a failure.) But our conclusion is an assertion of (1) with a difference. The implicature (I) has been cancelled. In asserting '(1) is not true' we have committed ourselves to its truth: '(1) is not true' is true. Here we are applying to '(1) is not true' (i.e. to (1)) the predicate 'is true'. But now the extension of the predicate we are applying has shifted (broadened). This new evaluation has an associated truth-schema which *will* apply to (1) (interpreted as it is at its original occurrence). But this truth-schema is not a truth schema associated with the original occurrence of (1). Rather, it is truth schema associated with our reflective evaluation '(1) is true'. Substituting (1), as interpreted as that original occurrence, into *this* schema will not lead to contradiction. That is the bare bones gloss on how our theory handles the Liar paradox.

II

I now wish to state in brief compass the general principles that underlie the theory. The theory is couched within a formal language that is understood to model natural language. There are two sorts of principles in the theory: schematically stated *formal principles* governing the relations between extensions of different occurrences of 'true', and *pragmatic principles* governing how extensions of indexical occurrences are established in a context. I begin by stating the former.

The formal language L contains an infinite list of individual variables $y, y_1, y_2, \ldots$; an infinite list of variables over sequences: $\alpha, \alpha_1, \ldots$; an infinite list of variables over terms of L: $t, t_1, \ldots$; an infinite list of variables over variables of L: $x, x_1, \ldots$; an infinite list of variables over well-formed formulas of L: $\Gamma, \beta, \emptyset, \Gamma_1, \ldots$. We shall imagine that L has the resources to provide structural names, via Gödel numbering, of all its well-formed expressions. Rather than spell out this system, however, I shall indulge in the use of Quinean corners. L has the usual punctuation signs, and signs for negation, material implication, conjunction, disjunction, biconditionalization, universal quantification, and existential quantification. L has a finite (or infinite) list of non-semantical predicate constants which need not be specified except that it includes indentity and resources for expressing arithmetic and set theory. L has an infinite list of predicate constants formed by the following two operations: Either 'R' or 'Sat' may be followed by any cardinal numeral written as subscript:

$$R_1, \mathrm{Sat}_1, R_2, \mathrm{Sat}_2 \ldots.$$

(We shall call subscripts in this list *subscript numerals*.) Either 'R' or 'Sat' may be followed by 'i', 'j', or 'k', written as subscript, which in turn may be followed by the addition sign followed by any cardinal numeral:

$$R_i, \mathrm{Sat}_i, R_j, \mathrm{Sat}_j, \ldots, R_{i+1}, \mathrm{Sat}_{i+1}, \ldots.$$

(We shall call subscripts in this list *subscript letters*.)

Predicate constants generated by the first of these two operations represent particular *indexically* used occurrences of the natural language predicates '——— is rooted relative to sequence ═══' and '——— satisfies ═══'. 'Satisfies' is understood to bear the customary relation to 'is true'. $R_i(\Gamma, \alpha)$ could be eliminated in favor of $\mathrm{Sat}_i(\alpha, \Gamma) \vee \mathrm{Sat}_i(\alpha, \sim \Gamma)$, though I find it more perspicuous to retain it as a primitive. Thus these predicate constants in L represent uses, in a context, of natural language indexical predicates.

Predicate constants generated by the second of the two operations (the subscripting of letters) represent *schematic* occurrences of 'is rooted' and 'satisfies'. Except for the fact that such occurrences appear in the statement of our general principles, they will not be important to our present discussion.

On our view, there are no occurrences of 'true' in natural language that may be properly represented as lacking subscripts. For example, one cannot transcend L in natural langauge to get a subscriptless predicate 'true in L'. Any occurrence of 'true in L' in natural language should be interpreted as bearing some implicit subscript on 'true'.

Well-formed expressions are defined by the usual recursive rules. I shall not bother to lay out these rules here.

The formal principles first provide a recursive definition of 'R_i'. We let '$\alpha_1(t)$' mean 'the assignment of α_1 to t'.

(0) $R_i(\Gamma, \alpha)$, where the largest subscript in $\Gamma < i$.
(We understand that if there are no subscripts in Γ, 'the largest subscript in Γ' will denote 0.)

(1) $R_i(\Gamma, \alpha) \rightarrow R_i(\ulcorner \mathrm{Sat}_i(t, t_1) \urcorner, \alpha_1)$.
$R_i(\Gamma, \alpha) \rightarrow R_i(\ulcorner R_i(t_1, t) \urcorner, \alpha_1); \alpha_1(t) = \alpha \quad \alpha_1(t_1) = \Gamma$.

(2) $R_i(\Gamma, \alpha) \rightarrow R_i(\ulcorner \sim \Gamma \urcorner, \alpha)$.

(3) $[R_i(\beta, \alpha) \wedge R_i(\Gamma, \alpha)] \vee [\mathrm{Sat}_i(\alpha, \ulcorner \sim \beta \urcorner) \vee \mathrm{Sat}_i(\alpha, \Gamma)] \cdot \rightarrow$
$R_i(\ulcorner \beta \rightarrow \Gamma \urcorner, \alpha)$.

(4) $(\alpha_1) R_i(\emptyset, \alpha_1) \vee (\exists \alpha_1)(\alpha_1 \approx^x \alpha \wedge \mathrm{Sat}_1(\alpha_1, \ulcorner \sim \emptyset \urcorner)) \cdot \rightarrow$
$R_i(\ulcorner (x) \emptyset \urcorner, \alpha)$.

Where '$\alpha_1 \approx^x \alpha$' means '$\alpha_1$ differs from α at most in its assignment to variable x'. (This notation can, of course, be deabbreviated using only symbols of L, including the identity sign.)

Principles governing other logical constants are analogous.

(5) $R_i(\Gamma, \alpha)$ only if so determined by (0)–(4) or analogues.

(We state (5) in English for brevity. It could, of course, be stated within L using set theory and Frege's method for turning recursive definitions into explicit definations.)

The subscripts in (0)–(5) may take either subscript numerals (for indexical occurrences) or subscript letters (for schematic occurrences) as substituends. We shall be primarily interested in substitutions of subscript numerals.

The intuitive idea behind these principles is close to Tarski's language-levels idea, though there are two modifications.[2] In the first place we do not treat sentences that lead to paradox as ungrammatical in the formal model, but rather count them unrooted$_i$ (for some i). Subsequent principles (esp. (6)) place semantical instead of syntactical strictures on such sentences. The reason for this modification of Tarski's idea is that sentences in English that lead to paradox cannot in general be recognized by mere syntactical criteria. A sentence may be self-referential, and may be a source of paradox, because of empirical facts, as Tarski's own discussion illustrates.

These facts may even be unobvious or unknown.

The second modification constitutes a liberalization of Tarski's 'level' strictures. The analogue to Tarski's strictures would be to convert the schemes in (0) into biconditionals: Any sentence containing 'satisfies$_k$', $k \geqslant i$ would be counted unrooted$_i$ (ungrammatical for Tarski). The intuitive idea behind the liberalization is that many sentences that contain semantical predicates with given subscript i do not lead to trouble when applied to sentences containing semantical predicates carrying j for $j \geqslant i$. For example, '2 + 2 = 4 or this very disjunction is not true$_2$' is seemingly true$_2$ (and true$_1$). 'Something Nixon said is not true$_2$' is true$_2$ regardless of whether Nixon said things involving 'true$_2$', 'true$_3$', etc. In these and similar cases, the evaluation of the sentences as true$_2$ or not true$_2$ is fixed by components of the sentence, or instances of its quantification, which either are non-semantical or involve semantical predications with lower subscripts. (Cf. principles (3) and (4).) The first sentence is made true$_2$ (and true$_1$) by its first disjunct. The second is made true$_2$ (and true$_1$) by Nixon's various non-semantical political prevarications. By contrast, '2 + 2 = 5 or this disjunct is not true$_2$' is rootless$_2$ because the first disjunct does not fix the truth$_2$ or untruth$_2$ of the sentence, and the second clearly leads to trouble. On the other hand, the same sentence is rooted$_3$ because its second disjunct is true$_3$.

Another set of cases which would violate Tarski's strictures, but which seem innocent of mischief, involve iteration. For example, '"Snow is white" is true$_1$' can be counted true$_1$ without harm. (Cf. principle (1).) Here the innermost occurrence of 'true' is applied to a sentence that is non-semantical and which fixes subsequent semantical evaluation unproblematically.

The remaining formal principles are relatively predictable:

(6) $\sim R_i(\Gamma, \alpha) \rightarrow \sim \mathrm{Sat}_i(\alpha, \Gamma)$.

All rootless$_i$ formulas are unsatisfied$_i$. So a sentence and its negation – cf. (2) – may be unsatisfied$_i$, though one or the other will satisfied$_j$ for some $j > i$.

Then we have, for rooted$_i$ sentences, the usual recursive characterization of truth:

(7) $R_i(\ulcorner \sim \beta \urcorner, \alpha) \rightarrow \cdot \quad \mathrm{Sat}_i(\alpha, \ulcorner \sim \beta \urcorner) \equiv \sim \mathrm{Sat}_i(\alpha, \beta)$.

(8) $R_i(\ulcorner \beta \rightarrow \Gamma \urcorner, \alpha) \rightarrow : \mathrm{Sat}_i(\alpha, \ulcorner \beta \rightarrow \Gamma \urcorner) \equiv .$
$\mathrm{Sat}_i(\alpha, \beta) \rightarrow \mathrm{Sat}_i(\alpha, \Gamma)$

(9) $R_i(\ulcorner(x)\,\emptyset\urcorner, \alpha) \rightarrow .\ \mathrm{Sat}_i(\alpha, \ulcorner(x)\,\emptyset\urcorner) \equiv$
$(\alpha_1)(\alpha_1 \approx^x \alpha \rightarrow \mathrm{Sat}_i(\alpha_1, \emptyset))$

This recursive characterization could be converted into a schema for explicit definitions of numerically subscripted truth predicates – in the manner of (5). But I forego doing so.

Satisfaction principles for each constant non-semantical predicate of L are included in the theory, together with this principle, which has an analogue for $\ulcorner R_i \urcorner$:

$$R_i(\ulcorner \mathrm{Sat}_j(t, t_1)\urcorner,) \rightarrow \cdot\ \mathrm{Sat}_i(\alpha, \ulcorner \mathrm{Sat}_j(t, t_1)\urcorner) \equiv \mathrm{Sat}_j(\alpha(t), \alpha(t_1)),$$

where '$\alpha(t)$' means 'the assignment of α to t'. Here again translation into L is trivial.

The restricted Tarskian truth schemas are:

(T) $R_i(S, \alpha) \rightarrow \cdot\ \mathrm{Sat}_i(\alpha, S) \equiv p,$

where 'S' stands for the name of any well-formed sentence of L and 'p' stands for the sentence itself.

Truth is cumulative:

(11) $\mathrm{Sat}_i(\alpha, \Gamma) \rightarrow \mathrm{Sat}_k(\alpha, \Gamma),\ k > i$

And iteration is governed by:

(12) $\mathrm{Sat}_i(\alpha, \Gamma) \rightarrow (\alpha_1)(\alpha_1(t) = \alpha \wedge \alpha_1(t_1) = \Gamma. \rightarrow$
$\mathrm{Sat}_k(\alpha_1, \ulcorner \mathrm{Sat}_i(t, t_1)\urcorner))$, where $k \geqslant i$.

These comprise the formal principles of the theory. In addition, there are pragmatic principles, which affect how numerical subscripts are established in context. We assume first a principle of *Minimalization* or *Beauty*:

(a) The subscript on occurrences of the predicate 'true' (or 'satisfies') is the lowest subscript compatible with the other pragmatic principles.

The primary motivation of (a) is that it simplifies the application of the formal principles.

Two other pragmatic principles are relevant. One is the principle of *Verity*:

(b) Subscripts on occurrences of 'true' (or 'satisfies') are assigned so as to maximize the applicability of truth schemas to sentences and minimize attributions of rootlessness.

The pragmatic justification of *Verity* is that it exempts ordinary discourse from pathologicality or paradox except where a speaker's intentions, conventional meaning of words, or empirical facts force the issue. Other things equal, a set of sentences should be interpreted as non-pathological (rooted).

A final pragmatic principle is that of *Justice*:

(c) Subscripts should not be assigned so as to count any given sentence substitutable in a truth schema instead of another, without some reason.

If A says 'What B says is not true' and B simultaneously says, 'What A says is not true' (and that is all they say), we are logically forced only to treat one of the statements as rootless. But since the two statements are in substantially the same relation to one another, neither is reasonably counted less pathological than the other.

Other pragmatic principles may be needed to connect the formal model to actual usage. But these provide a substantial start.

III

We now turn to applications of the theory to particular problematic cases. I will begin with a pair of two speaker cases, which I call 'tangles', and then proceed to some 'chain' cases.

I have discussed a two-person case raised by Kripke[3] elsewhere. But to illustrate the theory, I will discuss the example in somewhat greater detail. Suppose Dean says,

(i) All Nixon's utterances about Watergate are untrue;

and Nixon says,

(ii) Everything Dean utters about Watergate is untrue.

Each wishes to include the other's assertion within the scope of his own. By the principle of *Justice*, each occurrence of 'untrue' should be assigned the same subscript. To ensure *Verity* we assume that this subscript, i, is high enough to allow application of a truth: schema to any utterance by Dean or Nixon other than (i) or (ii) – and, by *Minimalization*, no higher. It is not necessary that one know what the subscript is. It is enough that this subscript be fixed by the pragmatic principles together with the empirical facts. The order in which the three pragmatic principles are applied is worth noting: (c), (b), (a). I think that this order is canonical.

Suppose Dean uttered at least one truth_i about Watergate. It follows by (4), (5), (12), and (9), that Nixon's (ii) is rooted_i and not true_i. If none of Nixon's utterances other than (ii) are true_i, then since (ii) is not true_i, Dean's (i) is true_i (by (4), (5), (12) and (9)). If Nixon did say something true_i about Watergate other than (ii), then Dean's (i) is rooted_i but untrue_i. All of this accords with Kripke's intuitions.

Suppose now that no utterance about Watergate other than (i) by Dean is true_i. If none of Nixon's utterances other than (ii) are true_i, then neither utterance is rooted_i (by (0), (4), and (5)) and both are vacuously untrue_i. Moreover, both are true_{i+1} – by (0), (4), (9). On the other hand, if at least one of Nixon's utterances other than (ii) is true_i, then Dean's utterance is rooted_i and untrue_i, by (4) and (9), and Nixon's utterance is rooted_i and true_i. These results seem intuitively sound.

We now turn to a two speaker case, rather similar to Kripke's, introduced into the literature a quarter century ago. Suppose the policeman says:

(iii) Anything the prisoner utters (interpreted in its context) is untrue,

and the prisoner says:

(iv) Something the policeman utters (interpreted in its context) is true.

Each utterance is intended to include the other in its domain – so the policeman and prisoner are nicely entangled. It may seem plausible that it follows as a matter of logic that:

(v) Something which the policeman utters is untrue and something which the prisoner utters is true.

The initial plausibility of (v) rests on a pair of *reductios* as follows. First, suppose everything the policeman utters is true. Then (iii) is true. Then (iv) must be untrue – as is everything else uttered by the prisoner. But if (iv) is untrue, everything the policeman utters is untrue, contrary to our initial supposition. Second, assume nothing the prisoner utters is true. Then (iv) is untrue. So everything the policeman utters is untrue; in particular (iii) is untrue. But then something the prisoner utters is true – contrary to our assumption. (v) is consistent; and we have concluded that its negation seems to lead to inconsistency (given the postulated utterances of (iii) and (iv)). So (v) seems to be a logical consequence of these postulations.

But this conclusion itself leads to an unacceptable conclusion. For we can prove on the basis of (v) – a purported logical consequence of the fact that the policeman uttered (iii) and the prisoner uttered (iv) – that either the policeman uttered something other than (iii) that is true, or the prisoner uttered something other than (iv) that is true. I take it that this result is entirely unacceptable. It is obvious that it is logically possible that the policeman and the prisoner utter nothing but (iii) and (iv) respectively. Thus we cannot infer (v) on the basis of the fact that policeman uttered (iii) and the prisoner uttered (iv). The inference goes wrong by making the assumption that one can apply the truth schema (call it the truth_i schema) associated with occurrences of 'true' or 'untrue' in (iii) and (iv) to (iii) and (iv) without insuring that those utterances are rooted_i. Whether they are rooted_i depends on empirical facts.

How does our theory deal with the example? Suppose first that neither the policeman nor the prisoner says anything other than (iii) and (iv) respectively. By *Justice*, they receive the same subscript. There is no way to apply their associated truth schema to both of them, so *Verity* cannot make that schema applicable. *Minimalization* assigns the subscript '1' to occurrences of 'true' and 'untrue' in (iii) and (iv). By (0), (1), (4), and (5), neither (iii) nor (iv) is rooted_1. So each is untrue_1 by (6). By (0) and (4) both are rooted_2. (iii) is true_2; (iv) is untrue_2 – by (9) and its analogue for existential quantification. Thus something the policeman utters, interpreted in its context (namely (iii), with 'untrue' interpreted 'untrue_1') is true_2. This is, however, not what the prisoner stated in uttering (iv), since his utterance's contextual interpretation involved 'true_1', not 'true_2'. The prisoner's utterance merely has the same *conventional meaning* as our remark that something the policeman utters, under its contextual interpretation, is true_2. Thus the prisoner uttered a sentence whose conventional meaning can make for true (true_2, true_3 ...) assertions – but not in the context of the prisoner's utterance.

Suppose now that the policeman uttered something other than (iii) that was true_1 and the prisoner uttered nothing but (iv). Then (iv) is rooted_1 and true_1, by the analogues of (4) and (9). And (iii) is rooted_1 and untrue_1 by (1), (4), and (9). This result captures intuitions which back the plausibility of (v). I shall leave it to the reader to construct cases in which the policeman utters only (iii), but the prisoner utters something true_1 (or untrue_1) other than (iv). Other cases are also constructible and may yield different subscripts as the result of applying the pragmatic principles. But the basic frame of our solution is illustrated by the two suppositions about the background for (iii) and (iv) that we have already considered. The results seem intuitively satisfying.

We now turn from tangles to chains. First consider the 'descending' chain:

(I) (1) : (2) is true; (2) : (3) is true; (3) : (4) is true; ... .

There is no reason to give some of these sentences truth conditions at the expense of others, so *Justice* seems to counsel assigning all occurrences of 'true' the same subscript.[5] There is no way to treat these occurrences as rooted relative to their associated truth schema. *Minimalization* leads to assignment of subscript '1'. Thus all the sentences in the chain are rootless_1, and hence untrue_1. They are rooted_2, but untrue_2.

To take another chain:

(II) (1′) (2′) is true or 2 + 2 = 4
(2′) (3′) is true or 2 + 2 = 4
⋮
(n′) (*n*′ + 1) is true or 2 + 2 = 4
⋮

Justice urges a single subscript. *Verity* and *Minimalization* urge maximizing application of truth schemas beginning with the lowest subscript, '1'. Since the second disjunct of each sentence (*n*) is true_1, the whole sentence (*n*) is rooted_1 and true_1 by the principle for disjunction analogous to principle (3). Since (2′) is true_1, the first disjunct of (1′) is true_1 by formal principles (1) and (12). The same reasoning applies all the way through: all the first disjuncts are true_1. This result seems intuitively correct.

A chain like

(III) (1′) The first disjunct of (2′) is true or 2 + 2 = 4
(2′) The first disjunct of (3′) is true or 2 + 2 = 4
⋮

is similar insofar as each sentence *n* is true_1. But the first disjuncts form a 'descending' chain. Each first disjunct is unrooted$_1$ and untrue$_1$.[6]

Consider now the following:

(IV) (1′) (2′) is true or (1′) is not true
(2′) (3′) is true or (2′) is not true
⋮
(n) (n + 1) is true or (n) is not true
⋮

Intuitively, it might be tempting to reason by mathematical induction as follows: Basis case: If (1′) is not true, then the second disjunct of (1′) is true; so (1′) is true. Induction step: Assume (n) is true. Then the second disjunct of (n) is not true. So the first disjunct of (n) is true, so (n + 1) is true.

How does our theory treat the case? *Justice* argues for equality of subscripts. Rootlessness is unavoidable. *Minimalization* urges assigning the lowest subscript, '1'. All occurrences of 'true' in the chain are rootless$_1$, by (0)–(3) and (5); each disjunct and every disjunction is untrue$_1$. Each sentence in the chain *is* true$_2$, because the second disjunct of each sentence is rooted$_2$ and true$_2$. (Each second disjunct says that the containing sentence is not true$_1$, and, as we have seen, the containing sentences are not true$_1$ because they are not rooted$_1$.) The first disjuncts are rooted$_2$ and untrue$_2$. But even seen from the viewpoint of applying 'true$_2$', there seems no way to salvage the reasoning of the mathematical induction.

The theory is thus committed to regarding the intuitive argument from mathematical induction as fallacious. Seen as treating 'true' as univocal throughout, the argument commits the mistake in both steps of attempting to apply to sentences in the chain the truth schema associated with occurrences of 'true' in those sentences.

But why should one accept the theory's verdict instead of the intuitive reasoning? In the intuitive reasoning, the basis case concludes that (1′) is true. Following this reasoning, one can ask wherein lies its truth? The second disjunct is not true since (1′) is true. So the first disjunct is true. It says (2′) is true. Wherein lies (2′)'s truth? (2′)'s second disjunct is not true, so its first must be. It says (3′) is true. Wherein lies (3′)'s truth? And so on. Clearly we have here a variant of the descending chain we considered earlier. But the descending chain is intuitively pathological. So contrary to initial appearances, the present chain is pathological also. Both chains are strung-out analogs of (a): (a) is not true.

A second argument against the intuitive application of mathematical induction is that precisely the same train of reasoning as applied to the following chain would suffice to establish that 2 + 2 = 5:

(V) (1′) 2 + 2 = 5 or (1′) is not true
(2′) 2 + 2 = 5 or (2′) is not true
⋮

The Basis case would be: If (1′) is not true, the second disjunct of (1′) is true; so (1′) is true. The Induction step would be: Suppose (n) is true. Then the second disjunct of (n) is not true. So the first disjunct of (n) is true; so (n + 1) is true. But since the first disjunct of (n) is 2 + 2 = 5, we will have argued, *ad absurdum*, that arithmetic is inconsistent.

The trouble with the induction argument as applied to both these latter chains is that using the associated truth schema, it infers the truth of a sentence from the inconsistency (again assuming the use of the associated truth schema) of the negation of the sentence. But this mode of inference is perilous when semantical pathology is at issue.

One must consider not only the consequences of counting a given sentence true, but also the consequences of treating its negation true. Our vague intuitive question 'wherein does its truth consist?' reflects an intuitive demand which our theory places on true statements. Their truth must be appropriately derivative either from non-semantical sentences, or from sentences whose occurrences of semantical predicates have narrower extensions. That is, their semantical value must be fixed independently of the semantical evaluation; it cannot depend on the outcome of that evaluation. In Chain (II) we can argue from the truth of a non-semantical disjunct to the truth of each step. We argued from the conclusion that the Liar sentence is not true_1 to its truth_2 since intuitively it says that it is not true_1 and it isn't. The application of 'true_2' here is derivative from the failure of (rooted_1) application of 'true_1'. In chains (III)–(V), there is no comparable way of seeing the truth of a given sentence in the chain as derivative from an independently fulfilled truth condition. Similarly, in the policeman-prisoner example, on the assumption that neither uttered anything besides (iii) and (iv), there is no basis for seeing either utterance as consisting of a derivative application of the truth predicate.

It is this intuitive schematic notion of derivativeness that our formal notion *rootedness* is intended to capture. So far, it appears that our theory captures this notion without undue strictures on intuitively valid informal arguments. The theory also has the merit of bringing out defects in reasoning that may initially seem sound. But further investigation is clearly warranted.

University of California, Los Angeles

NOTES

[1] Tyler Burge, 'Semantical paradox', The Journal of Philosophy LXXVI (1979), pp. 169–198. In the present paper I will be employing Construction C3, ignoring other constructions developed in the earlier paper.
[2] Alfred Tarski, 'The concept of truth in formalized languages', in Logic, Semantics Metamathematics, J. H. Woodger trans. (New York, Oxford, 1956).
[3] 'Semantical paradox', op. cit., pp. 194–5. Saul Kripke, 'Outline of a theory of truth', The Journal of Philosophy LXXII (1975), pp. 690–716.
[4] L. Jonathan Cohen, Journal of Symbolic Logic 22 (1957), pp. 225–232. Cohen's case was couched in indirect discourse and was intended to show that indirect discourse is not formalizable. I think Cohen was right in thinking that the example puts pressure on the formalization of indirect discourse, but wrong in drawing the sceptical conclusion. He is explicit about applying his reasoning only to indirect discourse. I will, however, discuss the example here in its bearing only on direct discourse.

The hierarchical theory of Charles Parsons, 'The liar paradox', Journal of Philosophical Logic 3 (1974), pp. 381–412, which is broadly congenial and from which I have learned much, does not deal with specific intuitions regarding situations like the Nixon–Dean case or such tangles as we are about to discuss. It merely blocks contradiction. (Cf. p. 405 for Parsons' remarks on two-person cases.) More generally, there are no specifics about how, in general, truth value is determined by the semantical or pragmatic roles of sentential parts. Parsons' view also differs in how it specifies the hierarchy. The difference derives from rather broad divergences in strategy and motivation that are not immediately relevant to matters at hand.
[5] If *Justice* did not apply, then we could give truth conditions to an arbitrary number of the sentences in the chain. Thus we could assign '4' to the 'true' in (1), '3' to the 'true' in (2), '2' to the 'true' in (3) and '1' to the remaining occurrences of 'true'. All the sentences would still be untrue$_i$ for every i, but we would be providing for more assignments of truth conditions by a sentence's associated truth schema in the spirit of *Verity*. There would be no unique maximization of such assignments in the case of the descending chain, but some might feel that the point of *Verity* is that more would be better than less. There would then be a certain prima facie tension between *Verity* and *Minimalization*. On the other hand, one could stipulate that since a *unique* maximization of truth conditions was impossible, the principle of *Minimalization* should collapse all subscripts to '1'. The problem with the descending chain, however, seems more illuminatingly diagnosed by the principle of *Justice*. There is a clear element of arbitrariness in favoring some of the sentences with truth-conditions over others. The diagnosis of the pathologicality of chains should be substantially the same as that of tangles.
[6] I am indebted to W. D. Hart for mentioning Chains III and IV, to Richmond Thomason for expositional suggestions, and to John Pollock for catching an error in an earlier formulation of principle (0). This earlier formulation also occurred in 'Semantical paradox', op. cit., p. 189. The present principle (0) should be seen as supplanting principle (0) in op. cit. The present principle (12) also supplants the like-numbered principle in the earlier paper.