

TRUTH, INTERPRETATION AND CONVENTION T

The present essay is an attempt at a critical understanding of the theory of truth and interpretation developed by Donald Davidson.¹ There are two distinct aims of this essay : (1) to bring out the supposed foundations of the new semantics and (2) to critically analyse the latter's claim to be an extension of Tarski's semantics. I shall argue that the new semantics is non-Tarskian in its aim and method.

I. Truth Theory and the New Semantics

An absolute theory of truth with an axiomatized structure is the main contention of the semantics developed by Davidson. It has two co-lateral theoretic frames : one which builds up an absolute theory of truth for natural languages modelled on the Tarski-requirement of Convention T,² and the other which constructs a theory of interpretation for natural languages modelled on the theory of truth itself. These two frames together constitute a frame of understanding natural languages with a finitist structure and a holistic method.

The theory of truth which is the basic component of the new semantics can be called the base-theory T_1 . It has the following structure :

- (T_1) Truth is a primitive concept³ of a semantical system for a language L. Therefore a truth theory provides a theoretical framework for understanding the concept of truth within axiomatized structure for a language L.

T_1 has the following two requirements :

- (1) T_1 has to be axiomatized, because it yields theorems of the form 'S is true iff p' for every sentence p of the language L of which S is the name in a metalanguage M.
- (2) The object language L must have a finite structure since a finite language alone can guarantee the needed axiomatization. The natural languages do have a finite structure which can be mapped up and learned.

The axiomatizability requirement is the basic demand of Davidson's semantic theory, since for Davidson a truth theory is an absolute theory, and not relative to any model or interpretation or possible world.⁴ A model-relative truth theory entails relativization of the concept of truth to the extent of making the latter a derivative concept. Davidson writes, "My thesis is only that there are important differences between theories of relative, and of absolute, truth and the differences make theories of the two sorts appropriate as, answers to different questions".⁵ Whereas a theory of relative truth argues for a truth-predicate applicable to a selected or accepted model, an absolute theory globally applies to all models of the classical sort which fit into the logic of the natural languages.

The key to axiomatization of T_1 is the availability of T-sentences which are generated from the Tarskian Convention T: 'S is true iff p'. The T-sentences which follow for every sentence of the language L can themselves be treated as axioms,⁶ and to that extent they are necessarily and globally true of the whole of the language L which, for Davidson, is none other than a natural language. Thus, the axiomatic availability of T-sentences is the *sine qua non* of an absolute theory of truth because by bringing out the intuitive strength of the concept of truth, the T-sentences pin it down as a primitive concept such

that no explicit definition is called for. This entails for Davidson two consequences :

- (1.1) Truth-definition fulfilling the Tarskian Convention T *need* not demand a syntactically rich metalanguage and so the syntactic resources of the metalanguage M need not be proliferated.
- (1.2) Truth-definition is recursive in the proposed semantic system without demanding an explicit character in terms of more primitive semantic concepts.

So far as the first consequence is concerned, Davidson's kind of semantic theory does not presuppose the distinction between formalized natural languages, and therefore the Tarski-requirement of richness⁷ of metalanguage for any formalized object language is not sought to be seen as a primary requirement. The natural languages which guarantee T-sentences have the required syntactic resources to carry on truth-definition. The second consequence, on the other hand, makes truth-definition easily accessible to all natural languages because the notion of 'true-in-L' is as much available for a natural language as for a formalized one. Admissibility of truth-definition in terms of satisfaction in no way dilutes the primitivity of the concept of truth and hence its recursive character.

In the light of the axiomatizability requirements the finitist demand, as mentioned above, is proposed as another demand of the natural language semantics. A finitist need is recognizably a syntactic need which is tagged onto the need to absolutize the concept of truth for mappable language. Davidson is aware of the syntactic availability of the finite vocabulary which provides the resources to build up infinite sentences of the language.⁸ Natural languages eminantly fill the bill sofar as the need of constructing infinite sentence is concerned. However, the finitist

constraint on the basic vocabulary is the built in constraint which is common to both natural language and formal language semantics. A language after all is intended to be mappable and so learnable to the extent it is endowed with a structure of rules.

An absolute theory of truth, as Davidson intends it to be, imposes finitude for the following reason :

- (2.1) Since T-sentences axiomatically extend over the whole of natural language L, there must be presupposed the constructibility of the infinite sentences. But the latter itself is the result of a finitely mappable language.

The above reason contains, in fact, two significant dimensions of the axiomatized truth theory. They are :

- (2.1a) Truth theory is holistic insofar as it entails T-sentences for all the infinite sentences of the language L.
- (2.1b) T-sentences are basically *entailed* by the internal syntactic resources and therefore those resources must be internalizable by the native speakers of the language L.

Both these dimensions are integrated into the truth theory, since in its holistic form it makes T-sentences available globally for the language L, and in its learnable aspect it puts truth theory closer to natural language, and thus makes it applicable to the latter. It is the latter fact which explains why Davidson builds up a semantics for natural languages which makes truth theory a central component.

II. Truth and Interpretation

The test of a truth theory, according to Davidson, lies in its empirical applicability⁹ to natural languages and thus in having the empirical constraint which complements the formal constraints already mentioned. The empirical constraint is of the nature of

interpretation depends not on the principle of charity but on what the principle tries to embed truth-holding in i.e., the rock-bottom beliefs. This completes the process of situating interpretation at the centre of interlinguistic communication which has been a central problem of natural language semantics.

As indicated earlier, both T_1 and T_2 constitute the frame-work of understanding of natural language. This involves working out a theory of meaning. But Davidson claims that once truth theory and the theory of interpretation are available the theory of meaning takes care of itself. That is, meaning is constituted by the truth-conditions¹² of the sentences as uttered by the speakers and as embedded in their belief-system. If the radically interpretable T -sentences are holistically available for the natural language, the availability of the meaning of the sentences of the language is necessarily guaranteed.

III. *Convention T and its Non-Tarskian Use*

Davidson's semantics, as brought out in I and II above, has made Tarski's Convention T its central concern. It is claimed to be the foundation of the finite axiomatizability of truth theory and its interpretative applicability to natural languages. Davidson writes: "Central to interpretation... is a theory of truth that satisfies Tarski's Convention T (modified in certain ways to apply to natural language). Such a theory may be taken as giving an interpretation of each sentence a speaker might utter."¹³ Given this thesis, it necessarily follows that Tarski's Convention T which is proposed originally as the material adequacy condition¹⁴ of any truth-definition is not only fulfilled by truth theory in the Tarskian sense but is taken beyond its formal limits to be embedded in an empirically testable theoretic structure. As far as its empirical embeddability is concerned, it is the resulting T -sentences which are having empirical parameters and are directly the vehicles of interpretation of sentences.

the "fittingness" to the natural language which a truth theory necessarily has. The theory which emerges out of the empiricalisation of the T_1 -sentences is called the theory of interpretation for a language L to be referred to as T_2 . We can formulate it as follows :

(T_2) T_1 -sentences like "snow is white" is true if and only if 'show is white' strictly represents what the native speakers of the language L hold as true, and thus represent what they believe and mean in uttering sentences of L .

T_2 , thus succinctly put, transforms T_1 into a theory of *interpretation* the native speaker's sentences so far as the latter fit into the truth theory. This fittingness is *empirically generated* through the maximization of the agreement in holding sentences true in the circumstances empirically given but remaining normal in not subverting the working of the system of beliefs of the native speakers. The following two constraints are necessarily imposed on T_2 :

($T_2.1$) The principle of charity¹⁰ works up the maximization of agreement to the extent of securing the slack between what is in fact true and what is believed to be so.

($T_2.2$) Theory of interpretation takes account of what the native speakers *actually* believe and mean in holding the sentences true, and thus the speakers' belief-system constitutes the bedrock on which the radical interpretation of the speakers' utterances stand.

The principle of charity is an empirical principle standing as a base-level principle for making interpretation, and so translation of the speakers' utterances possible. It extends the 'holding true' of sentences as far as practicable without eliminating the possibility of error completely.¹¹ However, the radicalisation of

As claimed by Tarski, Convention T entails T-sentences which are of the nature of biconditionals for all sentences of the object language L in a syntactically rich but finite-order metalanguage M. For Tarski, it is necessary that each of the infinite sentences of the language L has its T-sentence in the metalanguage M, and it is the test of adequacy of a truth-definition if the extension of the truth-predicate is secured throughout L. There are, however, two Tarskian constraints on Convention T :

- (Ta) T-sentences entailed by Convention T must be of finite order such that the former are of finite length and are embeddable in the metalanguage M which homophonically contains the object language L.
- (Tb) T-sentences are not axiomatized in M but are syntactically available in the hierarchy of languages.

The constraint Ta is of immense strategic importance since only finite-order languages can have an explicit truth-definition. If the object language has infinite-order expressions, then the metalanguage which includes the object language will be of infinite-order and thus will not be richer than the former and so has to contain its own truth predicate. In that case only an axiomatic treatment of truth is called for. Tarski writes, "... with respect to formalized languages of infinite order, the consistent and correct use of the concept of truth is rendered possible by including this concept in the system of primitive concepts of the metalanguage and determining its fundamental properties by means of the axiomatic method."¹⁵ The constraint Tb makes it obvious that T-sentences so far as they are available only in finite-order metalanguage are not axiomatized. There are no T-sentences in infinite-order languages and so no explicit truth-definition is possible therein.

Given the above Tarskian framework, it seems that Davidson's proposal to axiomatize the truth theory for finite-order languages where T-sentences are available is non-Tarskian. As remarked earlier, Davidson is eager to absolutize the concept of truth and hence he faces the need of axiomatization. But now following Davidson's proposal if truth is taken as a primitive concept which the axiomatic method demands, then the Tarskian need of explicit truth-definition ceases to be important and the resulting semantics gets a different orientation.

The Tarskian thrust of semantics is entirely formal, since the emphasis is on how to explicitly define truth for formalized languages of finite order in a syntactically richer metalanguage. Given this framework, truth-definition is attempted for the formalized language with the help of the semantical concept satisfaction so as to make explicit how the truth-predicate is ascribed to the sentences on the latter's being satisfied by infinite sequences of objects.¹⁶ The condition of satisfaction is laid down by Tarski as follows: "X is a correct (true) sentence in the individual domain a if and only if $X \in S$ and every infinite sequences of subclasses of a satisfies the given X in the individual domain a ". A Tarskian metalanguage, thus, ensures the material adequacy of the above truth-definition for the object language sentences within a chosen domain of objects, sofar as the definition is squarely based on Convention T. For Tarski though truth is defined for particular languages relative to individual domain of objects, the metatheoretic results apply to all languages of the finite order, and thus the theory of truth is based on globally available semantic foundations.

Davidson's proposal, which convincingly extends the Tarskian results to natural languages, nevertheless, makes departure from the strictly Tarskian demands of a truth theory. This consists; as Kripke¹⁸ notes, in making truth theory less formally demand-

ing so far as the metalanguage is concerned and in providing a finite axiomatizability for the theory. This results in the following modifications of the Tarskian programme.

- (T_{m1}) Truth theory is axiomatized in the sense that from a finite set of axioms T-sentences are derived as theorems. Thus, truth is taken as a primitive concept in the metalanguage.
- (T_{m2}) The resulting theorems i.e., the T-sentences are required to be empirically tested insofar as they are extendable to the natural language context.

The T_{m1} is a decisive point of departure from the available Tarskian requirements of a truth theory. The requirements of axiomatizability not only diminishes the scope of the Tarskian metatheory but also makes the richer metalanguage for truth-definition dispensable.¹⁹ Truth being accepted as a primitive concept, the need of higher-order expressions in the metalanguage is shown to be an extravagance. Besides, Tarski's distinction between finite and infinite order languages and the requirement of axiomatizing truth theory for the infinite order languages alone is not taken due note of in Davidson's programme. It is assumed all the while that though the truth theory is designed for the finite order languages, it nevertheless, has to be axiomatized. This may be viewed as a non-Tarskian result.

The T_{m2} , however, equally makes a dent in the Tarski-requirement of Convention T. Convention T which, according to Tarski, accounts for the material adequacy of truth-definition is a rigid requirement so far as the finite order languages are concerned since the T-sentences following from Convention T are necessarily true in the hierarchy of language and thus true sentences constitute a consistent and complete system²⁰ so far as they are deductively characterizable, i.e., can be decided to be true at a

fixed point in the metalanguage. However, T_{m2} makes the Tarskian Convention T open to the following criticisms :

- (C₁) Convention T and the resulting T-sentence cannot be extended to the natural languages unless the latter fulfil the minimum extensional requirement.
- (C₂) T-sentences may not be available, unless a richer metalanguage is ensured, and even if the required metalanguage is available, the T-sentences may turn out to be contingently true if the extended model of truth-definition is non-classical.

C₁ makes Davidson's extension of Tarski's Convention T vulnerable to theoretical difficulties since the natural languages do not admittedly have a formal structure, i.e. an extensional structure which is the paradigm structure of formalized languages. Davidson, however, intuitively accepts that natural languages are no less formal, but this assertion may be true only of fragments of natural languages. So far as natural languages entertain intensional structures involving belief and intentions, it remains doubtful if the Tarskian results on the Truth-definition will be valid at all.

C₂, however, closely follows the Tarskian injunction that T-sentences are not available for the natural languages, since the latter are inconsistent so far as they remain semantically closed. This makes it imperative for any truth-definition to be embedded in a higher order metalanguage. Since Davidson claims that a higher-order language than the natural language is not necessary and that truth-definition is still available for the natural languages, it can be inferred that for Davidson object language and metalanguage are coextensive. But, as Tarski warns us, this will lead to semantic paradoxes. The following passage

from Davidson confirms that Tarski's warning is not properly heeded to: "But if we understand our metalanguage, we are using a system of concepts and a language which is the one for which we *really* want a theory, for it is this richer system that is our natural one. And fortunately the richer system does not raise difficulties for a truth theory satisfying Convention T, for it is extensional."²³ The claim made here is to the effect that natural language, being a rich extensional system, can appear as both metalanguage and object language. This is symptomatic of semantic monism which makes Tarskian results invalid.

Besides, given the fact that all models of truth-definition may not be classical, i.e., two-valued and extensional, the rigid availability of T-sentences cannot be guaranteed.²³ Therefore, all extended models must ensure that if T-sentences are to be necessarily true, there cannot be any departure from the classical Tarskian models. Davidson's extended model has been claimed to be Tarskian in the core, but given the fact that there can be serious discrepancy between a formalized language and a natural one, it cannot be guaranteed that there will not be considerable weakening of the T-sentences in the Davidsonian model. Since the complete formalization of a natural language is still an open question, the scope of the above criticism remains open.

It is beyond doubt that Davidson models his semantical results on Tarski's Convention T, but the thrust of his results is non-Tarskian to the extent that the latter are intuitively wider in scope and empirical in motivation. The theory of interpretation which is derived from the theory of truth is sought to bring out the desired fit between the T-sentences and the linguistic behaviour of the community. Thus, the empirical constraints imposed

on truth theory do suggest that Davidson wants to achieve more than what Tarski's truth theory was designed to achieve.

IV. Empirical Constraints and Interpretative Truth Theory.

The empirical constraints which mark off Davidson's truth theory embed it in the natural continuum of speech behaviour. This embedding brings together two frameworks already referred to and builds up a holistic theory which not only shows the truth-structure of sentences but also the conditions under which those sentences are held true by the speech-community. Thus, the holistic requirement and the requirement of empiricalisation go together. The following may be treated as the Davidson-requirements :

- (DR₁) The formal structure of T-sentences must be embeddable in the speech-behaviour so that the former represents the internalized rules of holding sentences as true, e.g. 'Snow is white' is true iff snow is white.
- (DR₂) The theory of interpretation that naturalizes truth theory must be available for the whole of natural language since it completes the process of empirically embedding the truth-structures.

The requirement DR₁ brings about relativization of truth to the speaker-time,²⁴ nexus in the way of a deliberate design to produce T-sentences for all sentences including the ones involving indexicals. The speaker-time relativization is a part of the attempt to bring truth-structures closer to the speech-habits of the native speakers. Once this is secured, the formal constraints are relaxed so as to produce the interpretation structures for the sentences whose T-sentences are available. DR₂ is a constraint on interpretation rather than on truth. It shows that if we know the conditions under which the native speakers

hold sentences true, we can also know what they mean by those sentences as true.

Davidson's credit lies in putting DR₁ as the foremost constraint which makes truth theory the foundation of a theory of meaning and interpretation. A theory of meaning necessarily involves, according to Davidson, a method of interpretation, since meaning can be situated only in an interpersonal and inter-linguistic framework where the speech-behaviour of the native speakers is interpreted and understood by the interpreters according to some common parameters. The basic parameter is the maximal agreement on what is held as true by both the native and the alien speakers. Given this framework, one can map the truth-structures within a given language system onto its interpretative structures. That is to say, one can coalesce the truth-conditions of sentences in a language with the conditions of meaning of those sentences.

But does this 'grand design', in the words of J. A. Foster,²⁵ stand? Foster's objection to Davidson's scheme is that though the letter is formally an extensional framework, yet while being mapped onto an interpretative scheme it gets itself involved in intensional idioms.²⁶ Davidson's reply to this criticism has made it clear that if a theory of interpretation involves intensional concepts like belief and intention, it is not a sign of its weakness but rather of its strength,²⁷ since it corroborates the fact that a theory does not need to be exclusively extensional. That is to suggest that the resources of a theory cannot be restricted to the resources of what the theory is about. In this sense, it may be claimed that though the truth theory is extensional the theory of interpretation can afford to be intensional. But one can ask how can there be perfect fit or matching between the t-structures and i-structures of the sentences if the two theories so widely differ? Since the matching between t-structures and i-structures

is already taken for granted, what remains to be justified is the method of matching between the two theories. Davidson's way out of this problem is the holistic method of cementing the divergences between the extensional and intensional languages. This method succeeds to the extent there are mechanisms to correlate these two languages. Davidson champions holism on a larger scale to accommodate divergent tendencies in the theories of semantics.

But can holism ultimately dissolve the differences between a truth theory (T_1) and the theory of interpretation (T_2)? There are two ways of answering this question from the holistic point of view :

- (H₁) There are no two theories T_1 and T_2 ultimately since at a semantic level truth theory and the theory of interpretation collapse into each other. Therefore, the so-called intensional structure of I-theory is resolvable into the extensional structure of T-theory.
- (H₂) Even if T_1 and T_2 differ in their semantic resources, they do not differ in their theoretic content. Truth and its semantic characterization being the theoretic content of both, what I-theory adds to T-theory is the network of beliefs and intentions which are embedded in natural language. But what is added is only a matter of form and not of content.

H₁ is closer to Davidson's own answer to the above mentioned question, since T_1 and T_2 can be maintained to be the same theory at a deeper level. But it seems to rob the I-theory of its special significance. Besides, resolving the intensional language of T_2 into the extensional language of T_1 is a far too difficult task which has not been undertaken so far. In this respect, the solution offered by H₂ is of a more liberal character. The I-

theory is given a speciality of its own which is derivable from its contribution to the general structure of the T-theory as applied to natural languages. Davidson seems to recommend the innocuous character of the intensional language of T, insofar as the latter does not detract from the theoretic content of the truth theory. But the inner conflict between T₁ and T₂ can still be detected in the structural content of the theories. The difference in the form is no small difference.

Department of Philosophy,
University of Hyderabad,
HYDERABAD-500 134 (A. P.)

R. C. PRADHAN

NOTES

1. Donald Davidson; *Inquiries into Truth and Interpretation*, (Clarendon Press, Oxford, 1984).
2. See Davidson, 'In Defence of Convention T' *Inquiries into Truth and Interpretation*, pp. 65-75.
3. Davidson, 'Semantics for Natural Languages', *Inquiries into Truth and Interpretation*, pp. 55-64.
4. cf. Davidson, 'In Defence of Convention T', pp. 68-69.
5. *Ibid.* p. 69.
6. cf. Davidson, 'Semantics for Natural Languages' p. 56.
7. Tarski, 'The Concept of Truth in Formalized Languages' *Logic, Semantics and Metamathematics*, — trans, J. H. Woodger, (Clarendon Press, Oxford, 1956), pp. 152-278.
8. See Davidson, 'Semantics for Natural Languages', p. 56. And also his 'In Defence of Convention T', p. 70.
9. cf. Davidson, 'Radical Interpretation', *Inquiries into Truth and Interpretation*, pp. 125-139.

10. See, Davidson, 'Thought and Talk' *Inquiries into Truth and Interpretation*, pp. 155-70.
11. cf. *Ibid.*
12. See Davidson, 'Truth and Meaning' *Inquiries into Truth and Interpretation*, pp. 17-36.
13. Davidson 'Thought and Talk', p. 161.
14. See Tarski 'The Concept of Truth in Formalized Languages', pp. 152-65.
15. *Ibid.*, p. 266.
16. *Ibid.*, p. 195.
17. *Ibid.*, p. 200.
18. See Saul Kripke: "Is there a Problem about Substitutional Quantification?" in *Truth and Meaning* eds. Gareth Evans and John McDowell (Clarendon Press, Oxford, 1976), pp. 337-40.
19. cf. *Ibid.*
20. See Tarski, "The Concept of Truth in Formalized Languages", p. 198.
21. *Ibid*, pp. 132-65.
22. Davidson, 'In Defence of Convention T', p. 73.
23. Anil Gupta; 'Truth and Paradox', *The Journal Philosophical Logic*, II (1982) pp. 5-15.
24. See Davidson, 'Truth and Meaning', pp. 33-35.
25. cf. J. A. Foster 'Meaning and Truth Theory', in *Truth and Meaning* eds. Gareth Evans and John McDowell, p. 23
26. *Ibid.*, pp. 21-23.
27. See Davidson 'Reply to Foster'. *Inquiries into Truth and Interpretation* pp. 171-79.