

On Davidson's "Saying That"

TYLER BURGE

Natural discourse sentences in which surface substitution of coextensional expressions may not preserve truth-value – 'non-extensional contexts' – form the chief watershed for theories of logical form and content. On one side are the various intensional logics, deriving from Carnap and C. I. Lewis. These logics, where applied, represent such contexts as flouting the principle that extensionally equivalent, or co-denotational, expressions are inter-substitutable *salva veritate*. On the other side of the divide are the various strategies, deriving from Frege and Russell, that preserve this principle at the level of logical analysis by denying that apparent counterexamples involve genuine substitutions of extensionally equivalent expressions. 'Non-extensional context' is deemed a solecism, justified by convenience only. These latter strategies take one of two forms. Either they argue, as Russell did, that surface syntax of the relevant contexts is misleading in suggesting that the purported counterexamples utilize exchanges of expressions of the same logical type to yield variations in truth value of the relevant sentence. Or they argue, as did Frege, that it is mistaken to count the relevant expressions as having the same semantical values in the relevant contexts. Versions of the two strategies can be combined.

It may be that the intensional-logic and Frege-Russell approaches will come to seem more similar than they have in the past. But as of now, they offer the first momentous choice for a theory of the formal-semantical workings of natural language.

In his account of indirect discourse in 'On Saying That', Professor Davidson stands on the Frege-Russell side of the divide.¹ I stand there too. The various intensional logics have made profound contributions to the understanding of modality – and lesser, but significant, contributions to understanding tense, deontology, and so on. But their handling of propositional-attitude discourse has been inelegant or unilluminating. They have done little to connect meaning with understanding – particularly of non-sentential expressions – and thought

¹ Donald Davidson, 'On Saying That', *Synthese*, 19 (1968–9), pp. 130–46. Citations of page numbers will occur in the text.

with conceptual abilities. And in my view, they have failed to come to grips with detailed use of language that leads to paradoxes for 'intensional notions'.

There are more general considerations for favoring the traditional Frege-Russell line. First among them is a worry about how intensional logics treat the relation between notions like necessity, knowledge and belief, on one hand, and the notion of truth on the other. All of these notions have apparently similar grammatical features; but truth is treated as a predicate, whereas the others are treated as operators. There is a corresponding difference, on the standard treatments within intensional logic, between the logic and semantics of the truth-theoretic metalanguage and those of the intensional object language. These differences do not seem to me to have been given a satisfyingly deep motivation.

Davidson's strategy for maintaining the substitutivity of extensionally equivalent expressions is more similar to Russell's than to Frege's. The appearance of counterexample is attributed at least partly to misleading grammar, rather than purely to mistaken assignment of semantical values. Davidson's account differs from Russell's in where it lays the blame. The problem is said to be in the sentence divisions of superficial grammar, not (with an exception to be noted in section IV) in the characterization of the grammatical types of expressions. Davidson locates the expressions that appear to resist substitutivity in a different sentence from the one that attributes the indirect discourse. Since substitutivity principles in logic apply to sentences taken one at a time, the counterexamples are disarmed. This purely technical innovation is, of course, ingenious. But what is striking about it is that Davidson manages to embed it in an account that goes some way toward working.

Like Russell's very different view, Davidson's has long seemed to me important and provocative, but implausible. I think that our grip on the syntax of the relevant discourse is too fundamental and firm for us to be persuaded that we have been misled regarding how to categorize the singular terms or how to divide among the sentences. At any rate, I think that our grip on sentence division is much firmer than our grip on the semantics of 'intensional' contexts.

My purpose here is to discuss three more specific problems for Davidson's account. I think that I can solve one of these. One of the other two is widely known; but I do not see solutions to either. I doubt that they have satisfactory solutions from Davidson's point of view. All three seem to me to raise interesting general issues for theories of the logical form of 'non-extensional' contexts. After discussing these problems, I shall conclude with some brief remarks on the relative merits of the Frege and Russell approaches to indirect discourse and the propositional attitudes.

I Davidson's account

Davidson proposes to give an account of the logical form and formal semantics of indirect discourse sentences like 'Galileo said that the earth moves.' I shall sketch only the outlines of the account.

The main idea is that the logical form of the illustrative sentence is that of

Galileo said that. The earth moves.

The word 'that' is held to be a demonstrative used to refer to the utterance of the sentence 'the earth moves.'² Expressions that are coextensive with 'the earth' (or with 'moves') may be freely exchanged without altering the truth value of the containing sentence 'the earth moves.' Such exchanges are irrelevant to logical substitutions in 'Galileo said that' (indicating the relevant utterance) since they occur in a different sentence. 'Said' is a predicate with an argument place for a person, the sayer, and an argument place for an utterance event, picked out by a demonstrative.

Davidson holds that that is all there is to the account of the logical form of indirect discourse (p. 142). There is in addition an account of the use or meaning of indirect discourse 'said'. 'Galileo said that' is held to be a 'definitional abbreviation' (p. 142) of a longer expression:

- (A) $(\exists x)(\text{Galileo's utterance } x \text{ and that utterance of mine make us same-sayers})$

where we ignore tense, and where 'that utterance of mine' refers to the utterance of 'the earth moves' that follows the utterance of 'that'.

The remark about definitional abbreviation is uncongenial with the remark (p. 142) that 'sentences in indirect discourse ... wear their logical form on their sleeves [except for the period, in our example, between "that and "the earth moves"]' and with the statement that 'said' is a two-place predicate (p. 142). For if 'Galileo said that' really abbreviates the above-cited expression, it does not wear its logical form on its sleeve. The ordinary expression hides a quantification over Galileo's utterance; and although the logical form of (A) is not explicit, it clearly makes reference to me, the reporter (both in 'my' and in 'us'), and thus appears to require an argument place not only for Galileo and my utterance, but for 'me', the reporter. One could either make 'said' two-place and revise (A) to $(\exists x)(\text{Galileo's utterance } x, \text{ and that utterance (indicating an utterance) samesays with } x)$. Or one could retain reference to the reporter and provide an extra argument place in 'said'. There are further options. To provide a full account, one would have to settle this issue. But settling it will not be necessary for our purposes.

The notion of samesaying deserves comment. Davidson's idea is that the notion is an 'unanalyzed part' (p. 140) of the 'content' of 'said'. It is invoked, I think quite legitimately, to explicate the use of the predicate; and its complexity (saying the same thing) is not seen as complicating the predicate's logical form. To use indirect discourse, one must master the practice of samesaying. One must be able to use utterances that are relevantly synonymous with the utterances of the original speaker. The utterance to which 'that' is held to refer is a performance whose point is to mimic the content (not necessarily the mode) of the original speaker's utterance.

It might be noted that the account that Davidson's theory provides of the conditions under which the truth value of 'Galileo says that' might change exactly parallels the accounts of more traditional theories. If one were to

² Davidson characterizes the utterance as an act. I suppose, though Davidson is not ideally explicit, that this act is to be distinguished not only from the sentence type 'The earth moves' but also from the sentence token. The issue will not be relevant to our purposes as long as the referent of 'that' is not the sentence type. On that point Davidson is quite definite; *ibid.*, p. 142.

substitute for our utterance of 'the earth moves' another utterance that did not samesay that utterance, then 'Galileo said that' (where 'that' is taken to refer to the substituted utterance) might differ in truth value from the original token of 'Galileo said that' (where 'that' refers to our original utterance). In agreement with Frege's view, Davidson's holds that exactly those changes that preserve 'samesaying' in the expressions following 'that' are bound to preserve the truth value of the sentence in which the attribution is made. Davidson's theory differs primarily in its account of sentence division and in its assignment of semantical values.

Except for its peculiar claim about syntax, the theory has an appealing simplicity. It invokes a relatively simple logical form, a straightforward logic, an ordinary ontology, and a familiar and inevitable notion of synonymy that need be no more precise than the standards governing reports in indirect discourse.

II A problem about samesaying

The first of our problems concerns the relation between the hypothesized demonstrative 'that' and the relation of samesaying. The problem emerges by comparing ordinary reports that contain demonstratives in the sentence following 'that' with embedded reports of indirect discourse.

First, consider 'Galileo said that that moves' (where our second 'that' refers to the moon). Davidson's theory parses the sentence as

(B) Galileo said that. That moves.

For the report to be true, there must have been an utterance by Galileo that samesays my utterance of 'That moves.' What does samesaying consist in here? Clearly, Galileo would have to have uttered something in which he referred to the moon (perhaps demonstratively) and predicated something synonymous with 'moves' of the moon. In other words, in order to samesay with Galileo by using a demonstrative like 'that' in our content clause, we must preserve Galileo's reference in the subject of his sentence token and preserve Galileo's meaning in the predicate of his sentence token. I think that this point is intuitively obvious. It does not depend on interpreting 'reference' and 'meaning' in any technical or very precise way. So far so good.

Now, consider 'Galileo said that Copernicus said that the earth moves.' Davidson's theory parses the sentence

(C) Galileo said that. Copernicus said that. The earth moves.

The first occurrence of 'that' must refer either to the act of uttering 'Copernicus said that' or to the act of uttering 'Copernicus said that. The earth moves.' It will not matter which for our purposes. The second occurrence of 'that', the one in 'Copernicus said that', must refer to the act of uttering 'The earth move'.

We need some terminology here. Let the sentence token(s) produced by the utterance referred to by the first occurrence of 'that' in (C) be called 'Alpha'. Let the sentence token produced by the utterance referred to by the second occurrence of 'that' in (C) be termed 'Beta'. So Alpha is the relevant token of

'Copernicus said that. The earth moves' (or of 'Copernicus said that'); Beta is the relevant token of 'The earth moves.'

To report Galileo correctly, our token Alpha must samesay with some token of his (or must bear a relation for sentence tokens that is analogous to the samesaying relation between utterances). What does samesaying consist in here? Clearly our token Beta must samesay some token that Galileo produced in the course of attributing a statement to Copernicus. And our tokens 'Copernicus' and 'said' must attempt some match (presumably in meaning or, for the name, at least in reference) with tokens that Galileo produced. But what of the second 'that' in (C)? Clearly we cannot expect it to preserve the reference of any demonstrative or other expression that Galileo used. For our second 'that' refers to an utterance of Beta, and clearly Galileo made no reference to that utterance.

The problem then is this. There is a need to specify a notion of samesaying that constrains us to preserve the reference of the second 'that' in (B) and allows us not to preserve any reference in our use of the second 'that' in (C).

Substantially the same problem may be stated differently. There is a need to give an account of samesaying that enables one to treat the second 'that' in (C) as a demonstrative referring to an utterance, even though its use does not preserve reference under samesaying. On the hypothesis that the second 'that' in (C) refers to an utterance (or utterance token), it clearly differs in function from the second 'that' in (B). The second 'that' in (B) illustrates the most common function of demonstratives in that-clauses: They pick out the same objects that the original speaker (or thinker) made reference to. Their use *preserves reference* under samesaying. Can one articulate a coherent notion of samesaying that explains a different function for the second 'that' in (C), one that fails to preserve reference?

The threat here should be apparent. If some such explanation is not forthcoming, then a natural conclusion would be that *if* 'that' is a demonstrative in 'says that', it must refer to something (in unembedded as well as embedded contexts) that will be preserved under samesaying. This something cannot plausibly be an utterance or token of any kind. It must literally be something that might be said by different people in common. Some sort of abstraction lies at the end of this reasoning. The resulting theory would be different in philosophically significant ways from Davidson's.

Can this problem be solved? I think so. We need to find a non-*ad-hoc* notion of samesaying that will generate failures to preserve the references of expressions, here demonstratives. I think that we need look no farther than the ordinary notion of translation. Although it is normal for literal translation to preserve, as far as possible, the referents of expressions being translated, there are certain contexts in which it systematically avoids doing so. These contexts exhibit various varieties of or approximations to self-reference.

Clearly, if e_1 is a term in sentence s_1 and e_1 refers to s_1 (or to e_1), then a truth-value-preserving translation of s_1 into a sentence s_2 , and e_1 into a term e_2 contained in s_2 (where $s_1 \neq s_2$ and $e_1 \neq e_2$), will make e_2 refer to s_2 (or itself) if and only if e_2 does not refer to what e_1 refers to. Briefly, non-homophonic translation preserves self-reference if and only if it does not preserve reference.

Where self-reference is more important to the point of a discourse than what is referred to, self-reference is preserved at the expense of reference. Some-

times self-reference is more important. The point can be established by a wide range of cases independent of indirect discourse. For example, translations of expositions of diagonal arguments in metamathematics often preserve self-reference instead of reference.³ One can get a taste of the phenomenon by reflecting on a translation of '(3) is not a theorem,' where '(3)' refers to the sentence that contains it. One should imagine (3) embedded in a diagonal argument. A translation into Polish would provide a sentence that referred to itself, not to our sentence (3). Otherwise the validity of the surrounding self-referential argument would not be preserved without adding special premises.

The example from metamathematics is meant to forestall a claim that self-reference-preserving translation is non-literal. Standard translations of metamathematical treatises can hardly be dismissed as literary indulgences. Self-referential translations are sometimes superior to reference-preserving translations at preserving 'information value' or 'cognitive content', in intuitive, non-theoretical uses of those phrases. Coherence of the surrounding discourse and soundness of argument are also better preserved. A reference-preserving translation would be less intelligible to a native Pole, inasmuch as it would describe the self-reference without illustrating it. But the primary purpose of translation is to make foreign discourse intelligible to one who does not already understand that discourse.

A variety of other examples of self-reference-preserving translation could be adduced. But perhaps the point is already sufficiently clear for us to apply it to the case at hand. Our problem was that although the demonstrative 'that' must preserve reference under samesaying in (B), it must not in (C). If the second occurrence of 'that' in (C) could be associated with some phenomenon like self-reference and if this phenomenon could be seen as part of the point of the discourse – more important to its point than actual reference – then a natural notion of samesaying would generate and explain the difference in referential behavior between the 'thats' in (B) and (C).

Once this sketch of a solution is drawn, it becomes evident how to begin filling it in. The point, on Davidson's view, of uttering a sentence that gives the content of the original utterance is to reproduce in the language or context of the reporter's report a sentence that says the same as a sentence produced by the original speaker. We need but tie the reference to the utterance of the sentence following 'that' in 'says that' together with a reference to the language (or context, or utterance) of the introducing sentence to elicit the relevant sort of self-reference.

Thus 'Galileo said that the earth moves' could be represented

(A') $(\exists x)(\text{Galileo's utterance } x \text{ and that (next) utterance of mine, taken in the context of this very utterance, make us samesayers}).$ The earth moves.

³ The general point of these pages is developed in greater detail and with greater precision (without special reference to Davidson's theory) in my 'Self-Reference and Translation' in *Meaning and Translation*, (eds) Guenther and Guenther-Reutter, (Duckworth, London, 1978). Examples of non-reference-preserving translation from various fields are cited.

The embedded sentence, 'Galileo said that Copernicus said that the earth moves' goes:

- (C') $(\exists x)$ (Galileo's utterance x and that (next) utterance of mine, taken in the context of this very utterance, make us samesayers). $(\exists x)$ (Copernicus' utterance x and that next utterance of mine, taken in the context of this very utterance, make us samesayers). The earth moves.

Note that sameness of interpretative context (see section IV) is insured by the transitivity of identity as one proceeds to deeper embeddings. It is critical to the explanation that the languages or contexts-of-lexical-interpretation of the introducing and introduced utterances be connected, even identified. Otherwise, the element of self-reference in the introducing predicate would have no effect on interpreting the referent of 'that' in 'says that'. Lacking this effect, we have no suitably general explanation as to why 'that' fails to preserve its referent, in embedded occurrences, under samesaying.⁴

The point of indirect discourse might be fairly taken to be to introduce and produce an utterance that gives the content of the original speaker's utterance, *in the interpretative-grammatical context of the introducing utterance*. Self-reference within the introducing utterance to its own context is thus very much in the spirit of Davidson's proposal. Perhaps it is not crucial to explaining why samesaying does not preserve reference that one include an element of self-reference in the very logical form of indirect-discourse sentences. One might take the self-reference to be embedded in a general convention presupposed in the use of indirect discourse. The failure to preserve reference, however, does demand explanation. And I know of no other than that of making explicit the element of self-reference in the exemplar approach that Davidson proposes.

The need to explain failures of reference-preservation under translation or samesaying is not peculiar to Davidson's account. It arises for nearly every theory of intensional contexts that postulates context-bound or language-bound exemplars (whether concrete or abstract) – rather than propositions that are language and context invariant.

The problem that we raised for Davidson's account is a variant of the Church-Langford translation argument.⁵ This argument has been commonly

⁴ It seems to me that this connection between the contexts, or languages, of introducing and introduced utterances might help Davidson explain why it is *ungrammatical* to follow 'that' in 'says that' with non-English words. As the analysis stands, it is unclear why this should be so: If the introducing and introduced sentences were really separate, then no grammatical rules and nothing in Davidson's original analysis beyond pragmatic considerations of convenience would prohibit switching languages after 'that' in indirect discourse. But the prohibition is on a par in its normative status with other grammatical rules. By construing indirect-discourse 'says' in the manner of (A') and (C'), and by construing 'context' in such a way as to require introduced utterances to maintain the grammatical and semantical norms suggested by the introducing utterance, we have some account in Davidson's terms, of the grammatical rule. Since I do not think that there are separate sentences in ordinary indirect discourse, I shall not pursue the details of this matter.

⁵ The tradition of appealing to exemplars, which Davidson's account continues, derives from Carnap, Scheffler and Quine. Scheffler's theory is an exception to the rule I am stating, but it fails to provide a satisfactory account, partly for reasons Davidson gives (p. 139). Further discussion occurs in my 'Self-Reference and Translation', and 'The Content of Propositional Attitudes' *Nous* 14 (1980), pp. 53-8.

rejected for bad or confusedly stated reasons. It seems to raise a quite legitimate question for exemplar approaches about reference preservation. Why should reference not be preserved under translation? I think, however, that the question can always be answered by appeals to self-referential elements that are important for the intensional discourse being represented. The details of the answer will, of course, vary with the approach. But all exemplar approaches interpret intensional discourse in such a way as to make the appeal natural.⁶

The invocation of self-reference-preserving translation does more than check a threat from more traditional theories of intensional contexts. It illustrates one of the ways in which traditional approaches over-idealize translation practice. What one preserves under good translation is far harder to characterize than traditional appeals to meaning invariance or proposition invariance would suggest.

III *A problem about what is said*

Our second problem is widely appreciated. I raise it less because I want to develop it in detail than because it provides a useful foil for the third problem. The difficulty is that the entities that Davidson appeals to as the referents of 'that' in 'says that' do not seem appropriate for certain quantifications on indirect discourse or relevantly analogous discourse.

Seemingly, we may say that Galileo said something at 6 p.m. 3/15/1638 in the indirect-discourse sense of 'said', without being committed to there being an actual utterance token of ours that samesays Galileo's utterance. Even though we reporters may have the resources to report Galileo's utterance, we may never actually utter anything that samesays his sentence token.

The problems raised by mentalistic discourse are more varied and complex.⁷ Occurrent thoughts yield a problem analogous to the one just adumbrated. But

⁶ Davidson uses the Church-Langford argument against Carnap's analysis of propositional attitude sentences, arguing that the analysis 'Galileo spoke a sentence in his language that means what "the earth moves" means in English,' would not preserve reference under translation (p. 135). Davidson revises this analysis to 'Galileo spoke a sentence in his language that means what "the earth moves" means in this language.' Davidson argues against this revision also. He says that in a translation into French, the demonstrative reference (of 'this language') will shift 'automatically' from English to French; but he says that the quotational singular term that denotes the English sentence will be translated in such a way that its reference will remain unchanged. The result would be false (since 'the earth moves' means nothing in French), though the original sentence in English is true. This argument against the revision involves two mistakes. The reference of a demonstrative does not automatically shift under translation; it takes some special condition to motivate its doing so. Moreover, since translation should preserve truth value, a translation that preserved the reference of the quotation but shifted the reference of the demonstrative would simply be a bad one. A sympathetic development of Carnap's basic idea would motivate self-referential translation – in which, roughly, both the demonstrative and the quotation would shift reference. There are examples of translation within quotation marks in ordinary translation practice that are motivated by self-reference. (Cf. note 3.) Carnap does not himself develop his idea in this way. But appeal to self-reference and its effect on translation seems to be needed by, and natural for, both Davidson's approach and Carnap's.

⁷ A lesser difficulty is that the samesaying relation needs replacing by some relation more appropriate to mental states and events. The subject need not produce utterances for each attitude for us to 'samesay' with. I think that it can be argued that propositional-attitude discourse in natural

mental states force a further difficulty. When we quantify over what people believe, we seem to be saying that there *are* certain truths or falsehoods that they are committed to and that have logical structure and semantical attributes (e.g., truth, falsity). But there are surely beliefs and other attitudes that are never expressed or characterized by any actual utterances produced by the reporter, the subject, or anyone else.

There is a further problem along these lines for quantification on that-clause position in semantical and modal discourse. 'It is true that $2 + 3 = 5$ ' and 'It is necessary that $2 + 3 = 5$ ' are clearly syntactically analogous to indirect discourse. The truth and necessity predicates seem to have at least one fewer argument place than indirect-discourse 'said'. But sentences that contain them seem to demand an account of logical form that otherwise parallels the account for indirect discourse. It seems that there are truths of, say, arithmetic, that will never be uttered, written down, or otherwise instantiated. There are plausibly even truths that will never be believed, or otherwise associated with any propositional attitude. There need be no 'concrete' state of a person that is associated with such truths. Similar remarks apply to necessities, probabilities and so forth.

All of these problems of quantification arise because Davidson's account of what is said is nominalistic. It appeals only to concrete entities as truth bearers – entities that are localized in time (probably in space as well). In the general case, there do not appear to be concrete entities of the right sort (or perhaps even number) to account for quantifications onto that-clauses.

There are, of course, various ways of coming to grips with these difficulties. One might interpret the quantifications in a special way – for example by reference to one's *ability* to substitute names (or demonstratives) for the quantified variable, or by reference to possible concrete sentence tokens. These issues are, of course, complicated and in some respects rather technical. I shall not develop them here.

I believe, however, that neither interpretation in terms of possibility (or ability) is particularly plausible as an exclusive account of natural discourse. Quantification onto that-clauses does not seem markedly different from ordinary objectual quantification in non-intensional contexts. Moreover, pressing the notions of ability and possibility yields familiar difficulties: ability to substitute given what? – what one already knows or understands? Further, I believe that neither interpretation in terms of possibility has been shown to be capable of avoiding all commitment to non-concrete entities. In any case, any attempt to make these interpretations nominalistic would be more plausible as a *reduction* of the commitments of natural discourse than as an account of its actual *prima facie* commitments.

I doubt that Davidson's account can be extended in a natural and conservative way to deal with these difficulties. I think that the most plausible response to the problem is simply to regard natural discourse as quantifying over abstract objects in the relevant cases.

language does not make reference to inner sentence-like entities produced by the subject even in the case of mental events, much less in the case of mental states. But perhaps Davidson need only hold that some utterance of the reporter 'gives-the-content-of' the subject's mental state or event.

The pressure toward abstraction derives from the need to interpret quantification in discourse that yields knowledge. We are committed to objects needed to interpret such discourse. Perhaps the commitments to abstractions can be shown to be reducible. I am doubtful, but I shall not consider this question here. I do want to emphasize, however, that such reductions must fall out of reductions of the discourse that we have been considering to other discourse adequate to the same cognitive purposes. Metaphysical or epistemic preconceptions about ontology, for all their philosophical interest, have in themselves little cognitive payoff to back their prescriptions or proscriptions. Ontological positions should derive from, not dominate, the interpretation of cognitively successful discourse.⁸ Grand metaphysical or epistemic schemes (such as nominalism or global causal theories of knowledge) have not established themselves as cognitively successful discourse.

Davidson raises two objections to approaches to indirect discourse that would postulate abstract expression types as the denotations of content clauses in indirect discourse (pp. 134–7). Davidson calls these 'quotational approaches'. One concerns preservation of reference under translation. It can be passed over here, since it can be answered using the sort of considerations needed to sustain Davidson's own approach in the light of the problem discussed in the previous section. (Cf. note 6.)

Davidson's other objection concerns the need to fix an interpretation of the sentence type that is postulated by such approaches. Since words (word forms) may be understood in infinitely many different ways, we need to fix the relevant understanding or interpretation of the word forms. As means of fixing an interpretation, Davidson considers only specifying a language, say, English. And he marshalls Quine's objections in *Word and Object* to languages or, equivalently, propositions: the individuation conditions of propositions and languages are obscure; such questions of individuation are unconnected to indirect discourse; and postulating languages or propositions as entities is shown to be mistaken by the indeterminacy of translation (p. 138).

I think that these objections are not sound. But even taken as acceptable, the objections tell not against abstraction *per se* but against a particular means of fixing an interpretation.⁹

⁸ The position derives from Frege's methodology in *The Foundations of Arithmetic* and Quine's pragmatism in 'Two Dogmas of Empiricism' and *Word and Object*. Quine sometimes deviates from the method, it seems to me, in his preoccupation with physicalism. Davidson has expressed sympathy with such pragmatism in numerous places. I am therefore puzzled by the nominalistic aspect of his account of indirect discourse. The problems that we have discussed in this section are known to apply not just to his view but to previous nominalistic approaches. Yet I know of nothing in his writings that mitigates their seriousness.

⁹ I do not intend this last sentence as a criticism of Davidson since he does not claim or clearly presuppose the contrary. On the other hand, he does not consider other options for fixing an interpretation. It should be noted that on Quine's conception of indeterminacy the logical form as well as the interpretation is comprehensively indeterminate. I shall be ignoring this complication. Even relativity to a choice about interpreting form, however, is compatible with abstraction. One may see the logical form of a sentence as roughly analogous to the way Quine sees the number associated with a numeral. Any of several forms will do within a certain system of interpretation. I am sceptical of Quine's position here; but I cannot develop either the scepticism or the position's ramifications in this sketch.

It is important to distinguish the issue over whether the entities denoted by the noun-phrases following propositional-attitude verbs are abstract from issues over the nature of interpretation. The two overlap but are not the same. There are two primary disputed issues regarding the specification of an interpretation. One concerns the cognitive status or type of objectivity of linguistic interpretation: Quine's indeterminacy thesis bears on this issue. The other concerns the type of idealizations about language- and context-independence that are most fruitful in studying language. What is the relation between the use of language (human activity) and truth bearers? How far can one reasonably idealize beyond momentary contexts of use in giving an interpretation?

But abstraction is compatible with non-objectivist positions on the first issue and with nearly any reasonable position on the second. Thus to return to the 'quotational' approach Davidson criticizes: One need not say that Galileo said (indirect discourse) some *S*, taken as a sentence of English. One might, for example, say instead that Galileo said some *S*, taken as it would be if I (the reporter) were to use it now. Compatibly with taking *S* to be abstract, one might regard interpreting Galileo (or indeed oneself) as relative to some non-cognitive decision about analytical hypotheses, and one might interpret my would-be use in as context-dependent a manner as one pleases.

The counterfactual in this specification of the interpretation of *S* is no more obscure than what is required to know how to apply 'samesays'. We have to master some procedures for interpreting utterances beyond those actually at hand, in order to make use of the notion of samesaying. As in the case of Davidson's analysis, the specification of how the abstraction is to be interpreted might be regarded as implicit in the use of indirect discourse 'said' or the propositional attitude predicate. Alternatively, and I think preferably, it might be seen as part of a convention presupposed in the use of such predicates in such a way as not to affect their logical form. (See Davidson, note 3.)

This is not the place to discuss the issues regarding interpretation that I mentioned two and three paragraphs back. The main point is that the issues over the cognitive status of interpretation, and the appropriate level of idealization in specifying it, are separable from the issue of abstraction. Commitment to abstraction *per se* is not impugned by the sort of arguments that Davidson and Quine urge in favor of indeterminist, non-intensionalist, and relatively contextualist (or in Quine's case, individualist) positions on issues of interpretation and intentionality. Postulating abstractions does not in itself do much to advance (or retard) our understanding of linguistic interpretation. But it improves and simplifies semantical theory.

IV A problem about validity

The original and primary purpose of a theory of logical form is systematically to identify those formal structures in a language or discourse that mirror the deductive arguments that are formally valid. Such a theory will inevitably serve not only an account of validity, but also an account of truth conditions, whose interest and importance Davidson has done much to make manifest. Giving a

semantical theory of truth will sometimes lead to the identification of formal structures that do not necessarily mirror deductive relations but that enter into non-deductive reasoning. But a minimum requirement on a theory of logical form is that it assign structures to sentences by reference to which valid arguments that intuitively depend on form can be systematically explained.

What it is to be a *formally* valid argument becomes a difficult and subtle question when pressed. But the answer is traditionally bounded by two contrasts. In the first place, formally valid arguments contrast with valid arguments that depend essentially on the meaning of lexical items that are 'non-logical'. In recent times, what counts as a logical term has sometimes been regarded as almost purely a matter of what the logician decides to give a logic of (what to hold constant under model-theoretic reinterpretations). But traditionally, logical terms are those that are relevant to all subject matters. This characterization is vague (because of 'relevant' and 'subject matter') and leaves room for philosophical dispute over cases (e.g. truth, number, necessity). But it will serve our purposes.

In the second place, formally valid arguments contrast with arguments whose validity depends essentially on a context of application. Thus 'That is a man; so that is a man or a woman' – or 'I am tired and hungry; so I am hungry' – are valid – but only in a 'context' (context of application) in which the two occurrences of 'that' pick out the same object (or the two occurrences of 'I' and present tense pick out the same person and time). These cases could perhaps be assimilated to cases in which validity depends on the 'meaning' of the non-logical term 'that' ('I', 'am') – but only at the prohibitive cost of assimilating contextual reference to lexical meaning.

There are trivializations of this second boundary condition on formal validity. For example, since all formally valid arguments, construed as sequences of sentences, depend for their validity on some context – say, a broad linguistic context – in which at least some of their expressions receive their 'intended' interpretations, it might be held that there are no formally valid arguments, or that there is no genuine contrast between formal and context-dependent validity.

The attempted trivialization depends on running together two notions of context. There is what one might call the *context of interpretation*, the environment that one investigates in order to determine the grammar and lexical meaning of the expressions of a language. No expressions have meaning apart from a context in which they are related to various entities that they represent, to action, and to various other expressions. By varying the context of interpretation one may, trivially, vary the meaning of expressions. So any argument can be trivially rendered invalid if the context of interpretation is allowed to vary between premises and conclusion. Saying that the context of interpretation is to be held fixed may be seen as a way of enforcing the requirement that one avoid the fallacy of equivocation.

Even after a context of interpretation is fixed, thereby fixing the lexical interpretations of linguistic expressions, one has not yet fixed what one might call the *context of application*. Such contexts are more or less local and short-lived circumstances in which an expression is actually uttered or otherwise used. Even given a fixed lexical interpretation (or set of interpretations for ambiguous

expressions), the references of indexical expressions vary with contexts of application. And lacking a specific context of application, indexical expressions lack reference, though they do not lack lexical meaning. Arriving at a precise theoretical meaning for 'context of application' is a delicate matter that we need not pursue. Our purposes are served by the obvious point that the form and lexical meaning or interpretation of both indexicals and demonstratives do not suffice to fix their referents. Hence they do not suffice to explicate intuitions regarding the validity of arguments involving such expressions.

The notions of context interpretation and context of application must each be distinguished from the notion of a *context of evaluation*. These are possibly counterfactual circumstances (or circumstance-descriptions) by reference to which necessity or validity may be evaluated. Arguments are evaluated for validity by varying the context of evaluation. Formally valid arguments do not depend for their validity on fixing a context of application, though arguments involving 'context-dependent' devices do. Argument itself depends on fixing a context of interpretation, on avoiding equivocation.

Serious questions do indeed attend attempts to explicate in depth the contrast between the notions of context of interpretation and context of application. In particular, there are issues about the levels of idealization from contexts of application that are appropriate in arriving at a lexical and grammatical interpretation of linguistic expressions: What aspects of an individual and his environment enter into fixing an interpretation? How do these aspects relate to actual momentary uses?

For all the difficulty of these questions there remains an important distinction between formal validity and validity that depends on a particular context of application. The project of formal logic and that of providing logical forms for informal discourse are based on distinguishing form from immediate context as well as from non-logical meaning or interpretation. These projects table the large philosophical questions about the relation between interpretation and application, and isolate forms for their purposes that they presume are invariant across at least some particular, immediate contexts of application. This presumption has been vindicated many times over by the success and fruitfulness of logic, pure and applied.

To do and apply formal logic, then, we have to isolate forms that we can presume remain the *same* through the duration of an argument. Or on a nominalistic logical theory, we have to give proof-theoretic and semantic principles that hold for all 'replicas' of given tokens through the duration of arguments. Such forms are given constant meanings or interpretations. We can ignore *how* 'context-free' these meanings or interpretations of constants are – presuming only that variation of interpretation and form with contexts of application is not the norm. Forms are *individuated* partly with an eye to providing for stability of interpretation.

Demonstratives and indexicals are expressions whose interpretations do normally vary with particular, momentary contexts. A logic that operates on discourse that contains such expressions cannot be purely formal. It must relativize its principles to contexts of application.

All of this is background for our third problem for Davidson's theory. The problem is that the theory fails to capture the obvious formal validity of various

arguments. I will raise the problem in two stages: the theory fails to capture the formal character of certain validities; it also fails to capture the validity of certain validities.

To begin with, take the argument by repetition: 'Galileo said that the earth moves; so Galileo said that the earth moves.' It is, I think, obvious that this argument, and infinitely more arguments in indirect discourse like it, are valid. Moreover, they are formally valid if any arguments at all are.

Davidson's representation of the argument is

(D) Galileo said that. The earth moves.

So Galileo said that. The earth moves.

Each occurrence of 'that' picks out a different utterance: the utterance that follows it. I shall assume that the sentence tokens produced by the utterances are replicas of one another and that there is no equivocation in interpretation between them. Both utterances make the reporter and Galileo, samesayers.

On Davidson's account, the argument does not come out formally valid. This point holds even if one explicates the notion of formal validity within a generally nominalistic account of truth bearers. To give the semantics of Davidson's representation, we must explain what the two occurrences of 'that' refer to in their respective contexts of application. For insofar as 'that' is a demonstrative, its referents *ipso facto* depend on the particular contexts of its application, not purely on its form and lexical interpretation. So in explicating the argument, there must be a further story relating the different utterances of premise and conclusion – over and above the story that insures that they have the same form and that their lexical interpretation has not shifted. To explain the validity one would have to enter further premises about the relation between the two occurrences of the supposed demonstrative 'that', or an extra premise about its referents. In sum, the argument can never, on Davidson's theory, be treated as an instance of 'p; so p'. This is a problem because there is no intuitive basis for thinking that such a further story need be told.

The same problem can be developed from another perspective. Let us lay aside for the moment complexities introduced by indexicals or demonstratives within sentential clauses following 'says that'. So we concentrate on sentences, like, 'Galileo said (tense ignored) that $2 \times 2 = 4$ '. As long as the lexical interpretation of the sentence following 'that' remains unchanged (as long as utterances of the sentence samesay one another), one can shift utterances and tokens of that sentence *ad libitum* without discernibly affecting the truth conditions, logical potential, or meaning of the indirect discourse. So claiming that the noun phrases following 'says' refer to particular tokens, in these cases, does no work. And it undermines, or at least pointlessly complicates, an account of the form and validity of relevant arguments containing indirect discourse sentences.

This reasoning suggests two conclusions: first, the specification of the referents of noun phrases following indirect discourse verbs does not depend on a context of application but only on a context of interpretation. So those referents are not picked out by a demonstrative. Moreover, since the referents of such noun phrases do not normally vary with contexts of application, either they are constant, concrete paradigms or (more plausibly, I think) they are

abstract. The referents of such noun phrases can be specified formally – by fixing the form (and lexical interpretation) of the sentence following ‘that’ – and independently of the particular context of utterance. Since we *can* specify the referents formally, we should. For if we do not, we cannot explain arguments by repetition in indirect discourse by appeal to their logical form. Since such arguments’ validity is clearly formal (by the reasoning of the preceding paragraph) we should explain it by formal means.¹⁰

Davidson’s theory is subject to a further, more powerful and more simply stated objection. The theory not only fails to represent the formal character of valid arguments in indirect direct discourse. It fails to account for their validity.

To see this, note that the utterances referred to in the premise and conclusion of argument (D) are different events. Call the first event ‘Alpha’ and the second, ‘Beta’. The truth of premise and conclusion is explained in terms of the referents or extensions of their parts. Suppose now that the premise of (D), explained according to Davidson’s theory, is true. The conclusion we suppose is also true. But can we conceive of a possible circumstance (a context of evaluation) in which the premise is true and the conclusion is simultaneously not true? Clearly we can. We need only conceive, counterfactually, Alpha to exist while Beta does not. Since Alpha and Beta are, on Davidson’s theory, individual concrete events, there is no necessity that they co-occur. In such a case the premise of (D) would be true, while the conclusion would be either false or without truth value. Thus the argument on Davidson’s interpretation is not valid. But the English argument to be interpreted is valid.

As indicated earlier, the counterexample is compatible with assuming that Alpha and Beta samesay one another (or make their utterers samesayers). Depending on the metaphysics of utterances, there may be further counterexamples. For example, if it is not *necessary* that where they both exist, Alpha and Beta samesay one another, then there will be countermodels for (D) even where Alpha and Beta both exist. The countermodel of the preceding paragraph does not, however, depend on metaphysical delicacies. There is simply no ground for thinking that the two individual, dated, concrete utterances logically (or even ‘metaphysically’) must co-occur.¹¹

¹⁰ In the previous paragraph we laid aside cases where indexicals or demonstratives occur in sentences following ‘says that’. When they do occur, the particular utterance of the reporter does, or at least may, matter to an account of truth conditions or logical form. Does this fact help Davidson’s account? I do not think so. The semantical behavior of demonstratives in that-clauses is recognizably different from that of other expressions. They do not motivate the linguistic problems that make non-extensional contexts difficult and interesting. On any account their role in non-extensional discourse must be treated differently from that of expressions that resist surface substitutivity. Proper treatment of this matter would require a positive theory. The main features of such a theory are explained in my ‘Belief De Re’, *The Journal of Philosophy*, 74 (1977), pp. 338–62; and ‘Russell’s Problem and Intentional Identity’, in *Agent, Language, and the Structure of the World*, ed. Tomberlin (Hackett Publishing Co., Indianapolis, 1983).

¹¹ In conversation David Kaplan pointed out that on certain views about the necessity of origin, it is necessary that a given father exist if his son does. But, as Kaplan also noted, almost no one would hold that the son must exist if the father does. An analogous claim that Alpha and Beta must both exist if Alpha does would be even more outlandish because there *need be no causal relations between utterances among which valid implications hold*. This latter point tells against revisions of Davidson’s theory which would take the second ‘that’ in arguments by repetition to refer anaphorically back to the first.

Thus Davidson's theory fails to account for the validity of arguments by repetition in indirect discourse. Clearly, if the validity of such arguments cannot be accounted for, the theory can represent as valid no valid arguments that essentially involve more than one occurrence of a *that*-clause. I think that this consequence is unacceptable.

I have ignored the option of denying that arguments by repetition in indirect discourse are valid. This is the sort of move that someone with primarily ontological motivations might make. I take it that the deductive implication is intuitively obvious, and that ontological theory should serve such cognitive practices as intuitively valid inferences, rather than vice-versa. Denying the argument's validity would be reminiscent of earlier attempts to obscure structure to save ontological doctrine – for example, the orthographic accident theory of the structure of belief sentences, which Davidson has ably criticized.

The argument in this section has rested on claims about what a theory of logical form for natural discourse should capture, and how. I am aware that there are those, for example in the Wittgensteinian diaspora, who have maintained that such a theory is pointless or impossible: formal studies of language are said to be worthless; only radically 'context-dependent' explications of semantical intuitions are supposed viable; or only global theories of what meaning, interpretation, or understanding 'consists in' are of interest. I shall not discuss this recrudescence in detail here. The view seems to me to depend on ignorance of linguistics and its use of work on logical form. Still, some very general remarks are perhaps apropos.

Radical anti-formalist positions, to the effect that judgments about validity and other semantical features can never be adequately captured by appeal to formal structures, often stem from reaction against the traditional tendency among logicians to underestimate the prevalence of such context-dependent or 'messy' phenomena as indexicality, vagueness, presupposition, implicature, metaphor, irony, malapropism, differences of idiolect or dialect, and so on. There is, however, a gulf between recognizing the complexities that these phenomena force upon theories of language and embracing the anti-formalist position. Numerous validity judgments that are widely shared generalize across arguments (and argument tokens). Many such generalizations parallel the recursive aspects of language mastery. To try to account for such phenomena 'contextually' or piecemeal (or to ignore them) would be to lose the strongest theoretical grip we now have on the semantical – and indeed, the pragmatic – aspects of language use. Worse, it flies in the face of well-established facts about grammar and its relation to logical consequence.

Theories of logical form depend on assuming some conception of form that generalizes, or idealizes, beyond particular momentary contexts of application. But as I suggested earlier in this section, they are compatible with numerous options regarding how the resulting structures depend on, and how they are to be attached to, contexts of application.

Often anti-formalist positions arise out of preoccupation with philosophical problems about semantical content or about what exactly theories of logical form are doing. The problems are genuine. Again, however, there is a gulf between recognizing them, or even being sceptical about their solvability, and the anti-formalist view.

Theories of logical form are compatible with a variety of philosophical views about what goes into interpreting a person's sentences, about the relation between a person's understanding and the forms of his or her sentences, and about what goes into interpreting those sentences. Considerable agreement on appropriate forms is compatible with unclarity or disagreement about the relationship between those forms and minds or behavior.

Here as elsewhere, success in a cognitive practice does not presuppose success in answering philosophical questions about it. Sometimes the philosophical issues persist; sometimes they dissolve or lose interest; sometimes some semblance of agreement is attained regarding them. Though often inseminated by philosophical questions, cognitive practices usually mature without depending on answers to them. Anti-formalist views are in the unenviable position of comprehensively denying, on 'philosophical' grounds, the viability of a cognitive practice, pursued by linguists, logicians and philosophers, that has already taken root.

V Fregean vs. Russellian approaches to 'intensional' contexts

We began by contrasting Fregean and Russellian strategies for explicating 'intensional' contexts. The Russellian strategy denies natural assumptions about grammar. The Fregean strategy denies common assumptions about the semantical values of sentential parts. Davidson offers one consideration against the general Fregean strategy. (I pass over specific arguments against specific versions of the strategy.) It is a consideration that has deflected many. Davidson writes:

Since Frege, philosophers have become hardened to the idea that content-sentences in talk about propositional attitudes may strangely refer to such entities as intensions, propositions, sentences, utterances, and inscriptions. What is strange is not the entities, which are all right in their place (if they have one), but the notion that ordinary words for planets, people, tables and hippopotami in indirect discourse may give up these pedestrian references for the exotica. If we could recover our pre-Fregean semantic innocence, I think it would seem to us plainly incredible that the words 'The earth moves', uttered after the words 'Galileo said that,' mean anything different, or refer to anything else, than is their wont when they come in other environments. No doubt their role in *oratio obliqua* is in some sense special; but that is another story. Language is the instrument it is because the same expression, with semantic features (meaning) unchanged, can serve countless purposes. (pp. 144-5)

Despite the wisdom expressed in its last sentence, I think that this passage encourages a common misconstrual of Frege's view based on too narrow a reading of his text. The reading is partly fostered in English speaking philosophy by too unqualified a translation of 'Bedeutung' as 'reference', and by the common (mis)construal of 'Sinn' as linguistic meaning.¹²

¹² Cf. my 'Frege on Extensions of Concepts, from 1884 to 1903', *The Philosophical Review*, 93 (1984), pp. 3-34, sections I and VI; 'Frege on Truth', in *Frege Synthesized*, Hintikka and Taiminea (eds) forthcoming.

Frege's strategy assigns semantical values to expressions in 'non-extensional' contexts that differ from their semantical values in other contexts. But there is no special need to think of the semantical values of terms in non-extensional contexts as their 'referents'. Nor can they be identified with meaning in the ordinary, intuitive sense of the word. There is a function from the semantical values of expressions in non-extensional contexts to their semantical values in ordinary contexts. These latter are plausibly identified with the expressions' referents. So, in Frege's theory, semantical values in non-extensional contexts uniquely fix referents. There is no reason why the Fregean strategy cannot grant, to anyone who insists on it, the intuition that terms in non-extensional occurrences retain their customary referents. What it must hold is that in such occurrences such referents do not enter into the computation of the truth values of the larger sentences. I think it unreasonable for anyone to have or rely on strong intuitions regarding this latter claim. It is a theoretical claim whose relation to intuition is quite indirect.

Bedeutungen of expressions other than independent sentences in Frege's theory are semantical values that are assigned in accord with the principle that a sentence's truth value is a function of the values of its parts. The *Bedeutungen* of singular terms in extensional contexts are referents. But it has been unfruitful in interpreting Frege to think of the *Bedeutungen* of sentences, expressions in non-extensional contexts, and even functional expressions, as 'referents'. The point of this remark is not to suggest that such *Bedeutungen* are ontologically neutral. The notion of *Bedeutung*, like that of semantical value, is a theoretical extension of the notion of reference. The point is rather that these notions are more theoretical than that of reference. There are solid theoretical reasons to think that the semantical values of terms in non-extensional contexts differ from those in extensional contexts. I think that intuitions about the reference of terms are of severely limited import to this issue.

This claim is associated with a methodological view about the status of various intuitive judgments in a theory of logical form. The basic evidence for such a theory (and for a theory of truth) derives from reflective judgments regarding the truth or falsity of sentences (or propositions) and the validity of arguments; from judgments about the reference of certain singular terms in certain unproblematic, extensional contexts; and from reflective judgments about the logically relevant grammatical structures of sentences. Perhaps needless to say, all evidence from these sources is defeasible and subject to theoretical reflection. But these sources yield agreement. Both our intuitive judgments on these matters and our theoretical understanding regarding the logically relevant grammar of sentences are more reliable and more definite than our intuitive judgments or theoretical understanding concerning the semantical values of sentential parts, with the exception of demonstratives and certain other singular terms in unproblematic extensional contexts. I think that this point is borne out by comparing the amount of agreement and knowledge there is regarding syntactical categories (in logic and linguistics) with that regarding the semantical values of non-sentential expressions (other than certain acknowledged singular terms, prototypically names and demonstratives). The point favors Frege's strategy for dealing with non-extensional contexts, though not necessarily his particular hypotheses, over Russell's. Frege's strategy builds from intuitive and theoretical strength.

The value of Davidson's theory, in my view, lies not merely in its challenges to the unpersuaded. It also lies in its shift of perspective. The theory takes over from Quine, and extends much further, the perspective on propositional attitudes that emphasizes interpretation. While I believe that this perspective has its pitfalls, it is several steps beyond traditional discussions that glibly appealed to metaphors of grasping propositions that are expressed by sentences. It has facilitated identification of philosophical problems about understanding and expression that underlie semantical theories of non-extensional contexts.