Quotation

Jonathan Bennett

from: *Noûs* 22 (1988), pp. 399–418.

In his paper 'Quotation', Donald Davidson contrasts three theories about how quotation marks do their work, that is, about how tokens like this one:

'sheep'

refer to the type of which the following is a token:

sheep

He rejects the 'proper name' and 'spelling' theories, and propounds and defends a new account of quotation which he calls the 'demonstrative theory'. I shall argue that the truth about how quotation works has points of resemblance with both the spelling and demonstrative theories, though it is not a mere combination of elements from those two. It is closer to Davidson's theory than to the others, and I have reached it by developing the pioneering start that he provided.

1. Ground rules for the discussion

A quotation is a physical particular—ink on a page, chalk on a board, etc.—comprising a pair of quotation marks flanking an inscription which I shall say is 'displayed' in the quotation. The displayed item is itself a particular, consisting of some ink on a particular page, chalk on a board, or whatever. It is not an inscription-type.

A quotation never refers to the particular inscription that is displayed in it. Normally, the immediate referent is some type of which that inscription is a token, as in the sentence:

The word 'sheep' has an unusual plural form.

When I write

At the end of his letter, he wrote 'Damn you!'

perhaps my quotation refers immediately to the token which he wrote. But even if that is so (and I am not sure that it is), a type is still involved—reference is being made if not *to* a type then *through* a type. For simplicity's sake I shall always speak of the type to which a quotation refers. When I speak of how a type of inscription is structured in space, or of how its parts are spatially inter-related, I mean to refer to the relations amongst the parts of every token of the type. Types, or universals, do not literally have spatial parts.

When I seem to compare a particular with a type, I always mean to compare the former with every token of the latter.

Throughout, I shall use the word 'expression' to mean *inscription type which is a minimal linguistic element or a linguistically structured aggregate of such elements*. An aggregate is 'linguistically structured' if its elements are spatially inter-related in some way that has significance in the language in question. Where the English language is concerned, an item is an expression if and only if it is a set of alphabetic letters, spaces and punctuation points, arranged on one or more horizontal lines. (So every expression belongs to a language, in a way, but it need not be syntactically proper or semantically significant.) I think that Davidson also uses the term in this way.

Quotation can be of items that are spoken rather than written, as when someone writes:

She shouted 'Help!'

And quotations can themselves be spoken rather than written,

with help from suitable pauses and intonation contours, or from more explicit devices such as 'quote-unquote' or the signaling of quotation marks with one's fingers. For brevity's sake I shall set these extras aside and focus strictly on written quotations of types viewed as having only written tokens.

When I want to refer to an expression, especially if it includes quotation marks, I shall often do so not by enclosing it in further quotation marks but rather by displaying it on a line of its own, as I did twice in my first paragraph. This is for elegance and perspicuity, and has no theoretical significance.

2. The simple name theory

Of the three theories that Davidson discusses, one says that quotations are (tokens of) unstructured proper names of the quoted expressions; you don't understand how a quotation refers to a given item by looking for a systematic relation between that item and any part of the quotation, e.g. what is displayed in it, any more than you understand how 'Spinoza' refers to Spinoza by looking for a systematic relation between that man and some part of his name. According to this 'simple proper name' theory, if what is referred to by a quotation is systematically related to what is displayed in it, that is a happy accident and plays no part in explaining how that quotation has that referent.

Although distinguished philosophers have seemed to commit themselves to this theory, I don't believe that any of them have meant to do so. In particular, it is not credible that Frege, Tarski and Quine were unaware that what is displayed in a quotation is systematically related to what it names.

Why, then, have they appeared to endorse the simple name theory? One reason is that they have been less concerned with the details of how quotation does work than with heading off some misunderstandings about it. Suppose that a misunderstander thinks a single word occurs in the same way in

> 'sheep' has five letters

as it does in

> New Zealand has lots of sheep,

so that any substitution for the word in one would be equally good in the other. This implies that

> 'the domestic animals that are our main source of wool' has five letters

is true, which it plainly isn't. To clear up this sort of thing, it is enough to distinguish the word from an expression referring to it, comparing that substitution with wanting to ride on the name of a horse, eat a picture of a steak, and so on. The remedial work doesn't require one to be precise about how exactly the quotation refers to the word; all that matters is that the word does not occur in the quotation as a semantic contributor to it as it does in the sentence about sheep in New Zealand.

What about Quine's propensity for saying that a quotation 'names' its referent? (Quine, 1960, p. 143; 1953, p. 140.) Well, some of us do use 'name' just to mean 'designate', and it seems reasonable so to interpret Quine, given the absurdity of the view that a quotation of a word relates to the word as 'Spinoza' does to Spinoza. I am glad to have his assurance that this understanding of his position is correct.

Still, although it is not needed to explain Quine's writings on quotation—or, presumably, Frege's or Tarski's—it may be worth noting that 'simple' proper names can have mnemonic features. The inhabitants of Horsefly are so named that a stranger could, unaided yet unmiraculously, pair each with his or her name once he had been told the names: there's Lefty, that's probably Bighead, I'll be surprised if that's not Dimples,. . . and so on. If that were the situation, each name of a villager would be spelled the same as some word that describes him; and someone could be sure that this was so because he knew enough of the names to be inductively confident about what the procedure was. Yet the names would still be semantically unstructured proper names.

If quotations related like that to the expressions they designate, they would be 'simple names' after all, without that requiring us to learn, one by one, what each quotation type refers to. But they are not. A quotation is linked with its referent by a rule, not a mere regularity. If someone proposes that the phrase 'the whole nine yards' be named by the expression

> 'the whole ball of wax'

this is not on a par with someone's proposing that this pretty, slim, blond, village girl be named 'Butch'. The latter wou1d be

infelicitous, out of line with common practice, tough on the girl, and so on; but the former would be a plain misuse of quotation marks, a flouting of the convention relating what is displayed in a quotation to what is named by it.

It might be replied that whereas the inhabitants of Horsefly merely have a custom or habit of bestowing names that are spelled like descriptions of their bearers, in nearby Tumbleweed there is a stern convention or rule requiring that this be done. If that is how matters stand with quotation marks, we can maintain both that a quotation is an simple name and that it is downright wrong to name an expression by a quotation in which a token of some other expression is displayed. I reply that that leaves no substantive content in the idea that the quotation is a name rather than a description: it grips onto the name's bearer through semantic conventions that take account of what the bearer is like, just as descriptions do; and the claim that it is a name spelled the same as a description, rather than that it is a description, is idle.

So the 'simple name' theory—which no-one has ever accepted—cannot be rescued.

3. The demonstrative theory

The other rejected theory will be examined later; first we should look at what Davidson affirms about how quotations succeed in referring. According to his 'demonstrative theory',

> 'sheep'

means about the same as

> the inscription-type instantiated here: sheep

or as

> The inscription-type of which a token appears In figure 1:

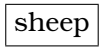

fig. 1

The strength and interest of this proposal can only be seen in the contrast between it and the spelling theory: Davidson claims at least four virtues for the demonstrative theory that its rival doesn't share. I shall come to them in due course.

4. The problem of relevant features

Any displayed token has countless features, and so is of countless different kinds. Therefore, to say

> the inscription-type instantiated here: sheep

or, what amounts to the same thing,

> the inscription-type each token of which is like this: sheep

is to leave things open to an intolerable degree. How to narrow it down? That is what I call the problem of relevant features. It urgently confronts the demonstrative theory, which must be amplified so as to meet it.

Here is an example of the difficulty. Suppose that in a hand-written document I write

> At the end of his letter John wrote 'Damn you!'.

If the quoted words are neat (or clumsy) in my quotation, do I imply that he wrote them neatly (or clumsily)? What about the relative sizes of the two words? What about printing versus cursive script? Or, if the whole episode occurs in print, roman versus italic? Gill sans-serif versus Garamond?

No. We do not regard our quotations as systematically committing us on details like those. If in a letter to you I quote John's curse, using heavier, thicker writing for the displayed words than for the rest of my letter, I probably intend to convey that John wrote those words more heavily than the rest of his letter, and I may well get that across. But that is not because quotation marks conventionally refer to, among other features, the thickness of the lines in the displayed item. Rather, it is for pragmatic reasons of the general kind that Grice has taught us about in his work on conversational implicature.

So the expression referred to by a quotation need not be exactly like the displayed item; but it is not enough that the two have some feature in common. The truth lies in between, and presumably has the form:

> A quotation in which x is displayed refers to an expression that resembles x in every respect of kind K,

with some value of K that yields a position intermediate between the too strong '. . . in every respect' and the too weak '. . . in some respect'.

5. Context *v* convention

The problem of selecting relevant features comes up with any illustrative procedure. If I tell you 'When he saw that it was Matilda, he scrunched his mouth up like this' and then I scrunch my mouth up, there is a co-ordination problem involving relevant features: what did I intend, and what did you understand, about how far and in what ways my grimace resembled the one it was illustrating? In many cases, including that one, the problem is solved—K is fixed—by general pragmatic rules together with the facts of the particular context. There are no general rules about illustrations of grimaces, but you and I will understand that my grimace is sufficiently like the original one, resembling it in the ways and to an extent that is suitable to our present needs and interests, whatever those might be.

'Show me how you moved your arm.' The illustrative standards that apply here might differ depending on whether the question comes from a chiropractor tracing a pain, a judge examining an insult, or a choreographer refining a composition.

Where quotation is concerned, however, I don't think that K is fixed in that way. Some of the fixing, at least, seems to come from the conventional meaning of quotation marks. I submit that quotation marks have the effect of fixing K at the value K = *linguistically significant*. I mean this to be confined to what is relevant to syntax and semantics, excluding anything that bears only upon typographic manner and the like.

If I am wrong about that, and K is fixed partly by the context, there will still be some conventional input from the meaning of quotation marks. That is enough to save my main argument in this paper, though it would require that some of the details be changed a little. From now on I shall ignore that possibility, and take it that convention and nothing else fixes K at linguistic significance and nothing else.

Where English is concerned, linguistic significance includes the respects that differentiate 'can' from 'con', but not those that differentiate a clumsily printed 'can' from an elegantly written one. There could be languages, however, where the difference between shapes like our printed 'a' and 'o' was a mere matter of handwriting, with no linguistic significance, while a shape like our handwritten 'a' represented a completely different letter.

Again, where English is concerned the linguistically significant facts would include ones about the directional ordering of letters on the line, but not about how many curve-containing letters occur on a single line. For all I know, the same holds for all actual writable languages, but it would be parochial and unphilosophic to build this in our account of how quotation works. For there could be languages where curve-occurrence was significant, or where words were determined by which letters they contained without reference to order; our present device of quotation would let us quote expressions drawn from such languages too; so a good theory of quotation will take account of them.

When Davidson writes that 'What appears in quotation marks is an inscription,. . . and what we need it for is to help refer to its shape' (p. 90), he invites the interpretation that the type referred to must be like the displayed inscription in every respect. But he moves on to speak of 'the expression of which this is a token' and 'the expression with the shape here pictured' (pp. 90f), and the replacement of 'inscription' by 'expression' makes things right again, implying that the referent shares with the displayed item all its features that are significant in the language concerned. At least, it does so if I am right in guessing that Davidson uses 'expression' as I do, to apply only to types defined purely by their linguistically significant features.

(Incidentally, Davidson's 'inscription' formulation seems to allow that we can quote items that don't belong to any language, for presumably any scrawl or sketch is an inscription with a shape. This seems to imply that it would be a proper use of quotation to write:

> At the very end of his letter, after his name but on the same line, he put ' ',

with a picture of a face, or a shapeless black blob, or a sketch of the Taj Mahal, between the quotation marks. If Davidson does allow that sort of thing, he has a problem about the selection of relevant features in those cases. But I doubt if that is his considered view, and I don't think it is correct. One might occasionally put a scrawl or doodle between quotation marks, and be understood to have referred to its shape, but only as a joking extension of

conventional quotation. Some recent advertisements for Meridien Hotels include a picture of a hotel servant, flanked by quotation marks—the caption is 'The French have a word for it'. There is clearly no literally correct way of understanding this *jeu d'esprit*, and certainly not as referring to the servant's shape.**)**

6. Enumeration *v* quantification

Now, a quotation might select the linguistically significant features of the displayed item in either of two ways, and corresponding to these there are two ways of enriching the demonstrative theory of quotation—one stronger than the other.

In (i) the strongly amplified theory, every quotation is tied to a given language L or kind of language K, and refers to the features of the displayed item that are in fact significant for L or for K. (By a 'kind' I mean a set of languages closed under the relation 'gives linguistic significance to exactly the same features as'.) Thus, every quotation means something of the form:

> The largest (weakest) type every token of which resembles the displayed item in the respects R_1, R_2,. . . , R_n,

and what the R's are will depend on what features of inscriptions are linguistically significant in the relevant language or language-kind. If the quotation purports to be quoting from (a language of the same kind as) English, its R's will involve the difference between 'x' and 't', and between 'o' and 'a', and so on; and also the difference between 'dog' and 'god' and so on. But they won't involve the difference between thick and thin lines, or between blue and black ink. A quotation purporting to quote from Martian, on the other hand, may include a very different list of R's, with emphasis on thick/thin rather than horizontal/vertical and on blue/black rather than left/right. The theory does not say that I understand a quotation only if I can recite the list of relevant features, for that demands too much. It says that I understand the quotation only if I have internalized the list, know what the relevant features are, can usually tell which pairs of inscriptions are and which are not tokens of a single expression in the language concerned.

In (ii) the weakly amplified theory, every quotation means something like

> The largest (weakest) type every token of which resembles the displayed item in all the respects that are significant in the language to which it belongs.

The difference between the two theories is that between a list of the relevant respects (strong) and the phrase 'the relevant respects' (weak). It is the difference between enumerating and quantifying.

Here is an argument for the weak theory. The strong theory implies that I cannot understand a quotation of an expression unless I know what structural principles are significant in the language from which it is drawn. For example, if someone tries to tell me on paper what the Martian word for 'planet' is, using a quotation in which the displayed material is a number of small inscriptions arranged on a wavy line, I do not understand this quotation (the strong theorist must say) because I do not know which features of the small inscriptions are linguistically significant in Martian, nor do I know what to make of the misalignments, whether left-to-right order is significant, and so on. My lack of understanding shows in my inability to replace the quotation by anything of the form

> The largest (weakest) type every token of which resembles the displayed item in the respects R_1, R_2,. . . R_n

because I don't know what the relevant R's are. Thus the strong theory. But I do understand the quotation: I do carry information away from it, namely that the Martian word for 'planet' is a type each token of which shares with that displayed item all of its linguistically significant features, whatever they may be. So the weak theory is right, and the strong one is wrong.

This has some force, but it can be turned into a pretty good argument going the other way, as follows.

If I don't know what features of inscriptions are significant in Martian, my grasp of any quotation of a Martian expression is essentially fragile. I can indeed extract information from it, but not in the way I do from quotations that I properly understand. Evidence for this: the only way I can safely pass the information along is by writing a perfect replica of the item my informant displayed in his quotation, like the tailor who when told to copy a suit reproduced the tear in its sleeve. In other words, I can convey the information only by passing along my whole basis for it; and that shows that I am operating on something I don't

really understand. This is the fumbling activity for which the weak theory is needed. Normal healthy quotation is nicely described by the strong theory.

7. Is there a right answer?

Let us try to decide between these two. Suppose that the very same token quotation of the Martian word for 'planet' is being contemplated by Kenneth who **k**nows and by David who **d**oesn't know what features of inscriptions are linguistically significant in Martian. Kenneth takes away from the quotation much more information than David does, as is shown by the fact that David can pass along what he has learned only by passing along the whole quotation or a perfect facsimile of it. According to the strong theory, that is because Kenneth fully understands the quotation while David does not. According to the weak theory, they are on a par in their understanding of the quotation, but Kenneth has an advantage in the further uses to which he can put the understanding, because of other things he knows.

This seems to be a stand-off; there may be no determinate right and wrong about it. The issue between the strong (enumerating) and the weak (quantifying) theories may be one of those cases that activates Quine's thesis that what a sentence means shades into what else would confidently believed by almost anyone who believed what it means.

Either way, something has to be added to the demonstrative theory to extricate it from the problem of relevant features. Throughout most of the remainder of this paper I shall run with the weak theory—not because I think it more plausible than the strong (indeed I find it slightly less so), but because I want to stay as close as I can to Davidson's theory, and the weak theory adds less to it than does the strong.

We have, then, a theory of quotation that is as close to Davidson's as we can get while surviving the problem of relevant features. It says that every quotation means something like this:

> The largest (weakest) type every token of which resembles this:. . . in all the respects that are significant in the language to which it belongs,

with a token of the referent in the gap.

8. The spelling theory

Now it is time to look at the other theory that Davidson rejects, which he calls 'the spelling theory'. According to this, the quotation

> 'sheep'

is tantamount to a definite description which means

> the expression you get if, moving from left to right, you write an 's' and then an 'h' and then an 'e' and then an 'e' and then a 'p'.

Davidson says that 'the appearance of quotation marks in the expanded notation is adventitious. . . Such a theory works as well, and is less misleading, if quotation marks are dropped entirely and new names of the [letters] are introduced' (p. 87). Thus, he prefers to take the spelling theory as saying that

> 'sheep'

is tantamount to a definite description which means

> the expression you get if, moving from left to right, you write an ess and then an aitch and then an ee and then another ee and then a pee.

To be equipped to go from any quotation to the corresponding description, Davidson says, one would have to 'know by heart the name for each smallest expression'. But if that is the whole load on one's memory, and if there are only twenty-six letters, that is easy. Clearly, the spelling theory improves on the simple name theory.

9. The 'new notation' objection

But Davidson charges it with four defects that the demonstrative theory lacks. Here is one of them:

> An important use for quotation in natural language is to introduce new notation by displaying it between quotation marks; this is impossible on the spelling theory provided the new notation is not composed of elements that have names. On the spelling theory we also could not use quotation to teach a foreign language based on a new alphabet or notation, for example Khmer or Chinese. Since these are functions easily performed by ordinary quotation. . . , we cannot accept the spelling theory as giving an adequate account of quotation in natural language. (p. 89)

I shall explain this through an example. According to our present version of the spelling theory, a writer of a logic textbook who says informally:

> I shall use '>' in such a way that anything of the form 'P > Q' means the corresponding thing of the form 'If it hadn't been the case that P, it wouldn't have been the case that Q'

means something which would be canonically expressed like this:

> I shall use corner in such a way that anything of the form
>
> pee ⌢ corner ⌢ queue
>
> means the corresponding thing of the form
>
> eye ⌢ eff ⌢ space ⌢ eye ⌢ tee ⌢ space. . .

and so on. And Davidson's objection (I take it) is that we readers to whom corner is being *introduced* won't already have a name for it, and so we won't be in a position to understand the informal statement as a colloquial expression of the canonical one; yet the writer will have succeeded in introducing corner to us; so it cannot be true that the latter of the above two statements is the canonical form of the former one.

The demonstrative theory, on the other hand, has no trouble allowing for our understanding of quotations of expressions drawn from radically unfamiliar languages. For we can understand phrases of the form

> The largest type every token of which resembles this:. . . in all the respects that are significant in the language to which it belongs,

irrespective of how unfamiliar the item in the gap is.

This argument assumes that the spelling theory would say that when you understand a quotation you must *already* have names for its minimal elements, and/or that you must give the same names that everyone else does; and that is questionable. Never mind. Davidson's next charge clearly requires the spelling theory to be amended, and that amendment will incidentally make it invulnerable to the new notation objection, even if the latter has all the force that Davidson credits it with.

10. The quantification objection

The spelling theory equates a quotation with a definite description containing names of letters, but the two are really quite different, Davidson says, because we can quantify into the latter but not into the former. From

> The expression you get if you write an ay and then an ess to the right of it is a word,

it follows that

> For some letter type x: the expression you get if you write an x and then an ess to the right of it is a word.

But from the true premise

> 'as' is a word

we cannot infer the meaningless or false conclusion

> For some letter type x: 'xs' is a word.

Davidson regards this as clear evidence that this version of the spelling theory 'is not a theory of how quotation works in natural language' (p.88).

In contrast, his demonstrative theory squares with this result and, indeed, explains it.

This is fatal to the spelling theory in the form in which Davidson has presented it, but before rejecting the spelling theory *in toto* let us see whether we can amend it so as to escape the quantification objection.

The latter depends on the fact that if a description contains a name, you can form a description from it by replacing the name with a variable and then binding that with a quantifier. Let us think about why names of letters were brought into the spelling theory in the first place. It was because the theory cannot be left in the form in which it associates a quotation with a description like this:

> what you get if you write an 's' followed by an 'h'. . . and so on.

This cannot be the basis for a complete account of what quotation marks mean, because it uses them in the explanans; so we need to modify the theory so as to get rid of those quotation marks, as Davidson says (I don't know why he says it is 'misleading' to retain them, but never mind that). He gets rid of them with the help of a loan from the simple name theory, replacing quotations of letters by names of letters. But we could get rid of them by borrowing instead from Davidson's demonstrative theory, replacing each expression consisting of a letter and a pair of quotation marks by

something of the form:

the letter that looks like this:

followed by a token of the letter. That would associate a quotation of the word 'sheep' with a definite description in which the minimal elements are referred to in this manner:

the letter shaped like this: s

the letter shaped like this: h. . .

and so on. The force of the word 'letter' is to make each of these equivalent to something like:

the largest (weakest) inscription-type every token of which shares with this:. . . every feature of it that is significant in the language to which it belongs,

with the displayed letter in the gap. That means that I am borrowing not from Davidson's initial demonstrative theory, but rather from the amplified version of it that was needed to cope with the problem of relevant features.

This form of the spelling theory is in no trouble from the quantification objection. It does equate each quotation with a definite description, but not with one in which the minimal elements are named. The falsity or senselessness of this:

For some letter type x: 'xs' is a word,

is exactly matched by the falsity or senselessness of this:

For some letter type x: what results if you write

an inscription shaped like this: x

then one shaped like this: s

is a word.

So the amended spelling theory is in no trouble from facts about quantifying in. And presumably I need not spend time explaining that it is safe against the new notation objection, however much force we allow that to have.

11. Reporting quotations: an aside

The amended spelling theory makes very explicit Davidson's view that quotation involves an indexical e1ement—a pointer or a 'this'. That has implications for what is going on when an *oratio obliqua* report includes a quotation.

If I report to you 'She said that I could keep *this*', holding up a photograph of her, I tell you nothing about how she referred to the photograph. My report may be true because she said 'You can keep *this*, brandishing the photograph, or because she said 'You can keep the photograph of me that I gave you', or because I said 'May I keep *this*?', pointing to the photograph, and she said 'Yes'. In the first of those cases she used 'this' as an indicator, but in my report it functions only as a quasi-indicator, signifying that she referred to the photograph *somehow*.

The terminology of 'indicator' and 'quasi-indicator', like my whole understanding of this phenomenon, is due to Hector-Neri Castañeda. For a recent discussion, and references to earlier ones see (Castañeda, 1981). So what my report says is something like this:

She said, of this photograph, that I could keep it,

which is conspicuously silent about how she referred to the photograph.

The same story applies if in a report I use not a 'this' but the quotation of a letter. If I tell you

She wrote that she always had trouble writing 'f',

I am saying

With regard to 'f', she wrote that she always had trouble writing it,

which could be true even if she didn't quote or otherwise display or indicate an 'f', but rather wrote:

I always have trouble writing the letter that looks like a Bishop's crozier with a cross-bar.

Someone who quotes a letter on his own account, uses an indicator; someone who quotes a letter in *oratio obliqua* uses a quasi-indicator.

That is for individual letters. If we move to larger units, as when I tell you:

She wrote that 'love' was her favorite word,

there is a slight complication. According to the amended spelling theory, if she wrote;

'love' is my favorite word,

she implicitly used four indicators, and one might infer that my report on what she wrote uses four quasi-indicators. But it doesn't. It bundles the four into one, thus:

Of the word 'love' she wrote that it is her favorite word,

or, if you want it spelled out further:

> Of the word that you get if you write a letter like this: l
> followed by one like this: o
> followed by. . . (etc.)

she said that it is her favorite word. The reason is that my report doesn't imply that she referred in any way to the separate letters of the word 'love'. The report would be true if what she had written was:

> The monosyllabic name of the supposed emotion of the heart is my favorite word.

The discussion in this section will be relevant, *mutatis mutandis*, to the remaining theories of quotation that I shall examine, for they both retain the Davidsonian indexical element. I shan't say any more about quotation in *oratio obliqua*; it is a routine matter to work out how the story should go from here on. Back, now, to main thread.

12. The picturing objection

The amended spelling theory, in the form in which I now have it, is less well armed against Davidson's charge that the spelling theory has 'no connection with the view that we understand quotations as picturing expressions' (p. 89). He assumes that when a quotation is in canonical form—expressed in a form that explicitly displays how it does it work—it will still depict its referent; and the complaint is that the spelling theory, unlike the demonstrative theory, does not guarantee this result.

Davidson discusses a prima facie objection to this, based on the idea that when you write out the spelling of a word you do depict certain aspects of it. His reply to this is right, in my opinion; but I shan't follow it out, nor shall I consider how Davidson might defend his premise that quotations essentially work by depicting their referents. Whatever we think about quotation and depiction, there is an independent reason why the spelling theory must be further amended; and when that has been done the resultant theory will imply that a quotation, canonically expressed, depicts its referent.

13. The parochialism objection

The spelling theory as Davidson and I have understood it is parochial. It has confined itself to languages in which the significant features are exhausted by **(i)** the facts about which types of inscription are linguistically minimal elements (letters, 'building blocks'), and **(ii)** the facts about how these elements can be arranged on a directional line. It is plausible to suppose that **(i)** is part of the story in every possible language, but there could easily be languages for which the rest of the story was not **(ii)** but something quite different, and we could construct and understand quotations from those languages as well.

This thought comes through a natural extension of a line of thought of Davidson's. He remarks that we can understand a quotation of an expression whose minimal elements are perfectly unfamiliar to us; I add that we could understand a quotation of an expression whose structural principles—ways of assembling its elements to form a complex whole—were equally strange. Although order on a directional line is the rule in all written human languages, there is no end to the possible alternatives:

> —There is linguistic significance in how the word-like clusters are ordered on the line, but within each cluster what matters is just how many tokens of which elements are present, their order within the cluster being a purely aesthetic matter.
>
> —Syntax is greatly affected by how large the space is between adjacent clusters.
>
> —The significance of a complex of clusters depends not only on their order on a horizontal line but also on certain vertical relations amongst the clusters.

And so on, ad infinitum. Confronted by a specimen of a radically unfamiliar written language, we shan't have *any* facts about its significant features handed to us on a plate. So we should not assume that when faced with a quotation from an unfamiliar language we even know which are the linguistically minimal elements of the displayed item, let alone knowing which of their features—and which aspects of their manner of composition—are linguistically significant.

For this reason, the notion of *spelling* is much too narrow. A properly general theory of quotation must not assume that when we understand a quotation from Martian, so to speak, we know what the elements are of the displayed item and know what facts about their manner of assemblage are significant. What, then, *do* we understand when we understand such a quotation? We understand the quotation to be equivalent to something like this:

> The expression which shares with this:. . . all its linguistically significant features,

with the displayed item in the gap.

That is my penultimately amended version of what began life as the spelling theory. Let us call it the *weak illustrated description theory* about what quotations mean—'weak' because it embodies the weak (quantifying) rather than the strong (listing) way of handling the linguistically relevant features.

That amendment was not motivated by Davidson's demand for a quotation that essentially depicts its referent, but it does incidentally meet that demand. For the weak illustrated description theory requires that a quotation in its canonical form, explicitly laid out in a manner that shows us how it works, still contains a token of the referent—not merely tokens of each of its minimal elements.

14. A point of convergence

Here is a slightly more prolix, slightly more explicit, version of my latest statement of what a quotation amounts to:

> The largest (weakest) inscription-type every token of which resembles this:. . . in all the respects that are significant in the language to which it belongs,

with the displayed item—a token of the referent—in the gap.

This is exactly the wording I used for my amended version of Davidson's demonstrative theory!

The weak illustrated description theory stands at the point where two lines of thought converge: one starts at the demonstrative theory and adds something descriptive to solve the problem about relevant features; the other starts at the spelling theory and adds something illustrative to meet one or more of Davidson's objections (new notation, quantification, picturing), while also modifying its descriptive part so as to meet my parochialism objection.

In adding to the demonstrative theory, and in revising the spelling theory, I have chosen descriptions that (weakly) quantify over the relevant features rather than (strongly) listing them. I shall reconsider that choice shortly.

The weak illustrated description theory differs from Davidson's only by deploying a notion of description that is not explicitly present in Davidson's account. It differs from the spelling theory in two big ways: it brings in a notion of illustration that is not to be found, even between the lines, in anyone's account of that theory; and it generalizes from the parochial notion of spelling. By any reasonable measure, it is closer to Davidson's than to the other theory, and for all I know it adds nothing to what Davidson intended when he advanced his demonstrative theory.

15. The semantic role objection

Still, it does add to what he actually said, and the addition needs to be made explicit. Only then can we arrive at a proper attitude to his fourth objection to the unamended spelling theory:

> A [theory of quotation should] provide an articulate semantic role for the devices of quotation. . . When we learn to understand quotation we learn a rule with endless applications: if you want to refer to an expression, you may do it by putting quotation marks around a token of the expression you want to mention. A satisfactory theory must somehow embody or explain this piece of lore. (p. 89)

Davidson charges that the spelling theory does not do this.

He doesn't actually say what 'articulate semantic role' his demonstrative theory assigns to the devices of quotation. The view about this that arises most naturally from what he writes is that quotation marks mean 'The inscription type instantiated here' accompanied by a pointer aimed at the displayed item; and that comes to grief because of the problem about the selection of relevant features. But when Davidson's theory has been properly amplified, making it identical with the weak illustrated description theory, it can still assign an articulate semantic role to quotation marks: each pair of them means 'the largest type every token of

which resembles this. . . in all the respects that are significant in the language to which it belongs'. That meaning for quotation marks, together with the differences in what is displayed in the gap, yields all the differences in what quotations refer to.

So the weak illustrated description theory can explain Davidson's 'piece of lore' that 'if you want to refer to an expression, you may do it by putting quotation marks around a token of the expression you want to mention'.

16. The strong illustrated description theory

I suspect that the weak form of the theory is wrong, however. As I said after first canvassing the weak/strong difference, I am inclined to prefer the strong theory according to which if I don't know what the relevant features of a displayed item are I don't really understand the quotation. Of course I can report:

> What I found on the outside of the package was something like this:. . .

followed by a more or less careful reproduction of what I found. But to report this in the form

> What I found on the outside of the package was ' '

with a more or less careful reproduction of it between the quotation marks—that strikes me as an abuse of the device of quotation. I am inclined to think that quotations should be regarded as addressed to, and understandable by, only people who know what the linguistically significant features of the displayed item are.

If that is right, then we should opt for the *strong illustrated description theory*, according to which every quotation means something of the form

> The largest (weakest) type every token of which resembles this:. . . in respects R_1,. . . ,R_n,

with the displayed item—a token of the referent—in the gap. This does just as well as the weak theory in meeting Davidson's quantification, picturing and semantic role objections, and also my parochialism objection. As for his new notation objection—it strengthens the case for rejecting that as not well founded.

That may sound more plausible about quotations from Khmer or Martian than about ones through which someone adds to our own language—for example introducing a new logical symbol. That is because when we are being invited to add something to our stock of English symbols, we bring to it our knowledge of which structural features determine the significant letter types in standard English, assuming that they will apply to the new symbol also, since we have not been warned otherwise. It looks a bit like a horseshoe on its side: we assume that we may write it a bit bigger, or a bit smaller, or in any color of ink, but not left-right reversed or rotated through 90° or with its one curve replaced by two right angles. These assumptions drastically reduce the extent to which the new symbol is unfamiliar; they domesticate it to the point where it is irrelevant to my present theme.

The strong theory lacks a convenient feature of the weak one. In the context of the latter, we could assign a semantic role to pairs of quotation marks by saying what they mean, that is, attributing to them a meaning that is also expressed by a certain English phrase. That cannot be done in the context of the strong theory. Where the weak theory says that quotation marks are synonymous with a phrase that includes the words 'all the respects that are significant in the language to which it belongs', the strong theory replaces quantification over significant features by a listing of them, and that makes trouble. If we declare quotation marks to be synonymous with some English phrase, either the phrase does or it doesn't include a list of linguistically significant features. If it does, what we say will be false because parochial: there will be possible languages whose significant features are omitted from the list. If it doesn't, what we say will not be consistent with the strong theory. The point is fairly obvious once it has been pointed out. (I completely overlooked it until Robert Van Gulick made it obvious to me.)

This would be bad news if it implied that a systematic account of how quotation marks do their work cannot be given consistently with the strong theory. But it doesn't imply that, and the account can be given. It goes as follows.

The semantic role of a pair of quotation marks is to express a *function* from inscriptions to descriptions based on them. In the following quotation, for example,

> 'sheep'

the function is expressed by the pair of quotation marks, its

argument is the displayed item, and its *value* for that argument is a definite description of the form

> The expression which shares with this:. . . all the features R_1, R_2,. . . , R_n,

with the displayed item in the gap, and with R's that are all and only the features that are linguistically significant in the language to which the displayed item belongs. The quotation is therefore semantically equivalent to that definite description and refers to whatever the latter refers to, just as '2^3' refers to the same number as '8'. It is not a defect in this account that it makes the phrase 'the word "sheep"' come out as 'the word the expression which. . . ' etc. Compare substituting 'the composer of *Parsifal*' for 'Richard Wagner' in the phrase 'the composer Richard Wagner'.

So the strong illustrated description theory can assign an articulate semantic role to pairs of quotation marks. It describes them as expressing a determinate function from token expressions to definite descriptions of expression types. That is quite enough to explain Davidson's 'piece of lore' that 'if you want to refer to an expression, you may do it by putting quotation marks around a token of the expression you want to mention'. It doesn't do so by equating quotation marks with an English phrase, but it is none the worse for that.

17. Another point of convergence

The strong illustrated description theory is also on a point of convergence. It is what you get if you start at the demonstrative theory and add to it, to solve the problem about relevant features, a descriptive element that involves (strongly) listing the relevant features rather than (weakly) quantifying over them; and it is also what you get if you start at the spelling theory, add something illustrative to it so as to meet Davidson's objections about quantification and/or picturing, and also modify its descriptive part so as to meet my parochialism objection, doing this in terms of (strong) listing rather than (weak) quantification.

Unlike the weak illustrated description theory, the strong cannot possibly be what Davidson intended all along: his new notation objection rules that out. But it is still a lot closer to Davidson's than to the spelling theory. I think it is the true theory about how quotation works, and I reached it on the basis of the flying start provided by Davidson.[1] –

References

Castañeda, Hector-Neri, 1981, 'The Semiotic Profile of Indexical (Experiential) Reference', *Synthese* 49, pp. 275–316.

Davidson, Donald, 1984, 'Quotation', in his *Inquiries into Truth and Interpretation* (Oxford University Press), pp. 79–92.

Quine, W. V., 1953, *From a Logical Point of View* (Harvard University Press).

Quine, W. V., 1960, *Word and Object* (M.IT. Press).

[1] This paper has profited greatly from Peter van Inwagen's and Mark Brown's help.