See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/266424601

Logical Forms: An Introduction to Philosophical Logic.

1 author:

Mark Sainsbury
University of Texas at Austin
60 PUBLICATIONS   **537** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Semantics of names View project

Philosophical Logic

> Instead of following grammar blindly,
> the logician ought rather to see his task as that of
> freeing us from the fetters of language,
> Frege (1897: 143)
>
> The old logic put thought in fetters,
> while the new logic gives it wings.
> Russell (1914: 68)

1. Introduction

The first use of the phrase "philosophical logic" known to me is in a semi-popular essay by Bertrand Russell called "Logic as the essence of philosophy" (1914). Russell and others had been practicing philosophical logic for years, but it was not until this essay that Russell identified and labelled what he considered to be a distinctive approach to philosophy as a whole. Philosophical logic, as Russell conceived it, was the programme of resolving traditional philosophical problems by discovering and classifying logical forms, and this concern with logical form became a central and distinctive feature of analytic philosophy in the twentieth century. Analytic philosophers do things other than philosophical logic; but those willing to regard their work as contributions to philosophical logic are analytic philosophers.

The most famous contribution to the programme was Russell's own theory of descriptions, whose first published version was given in "On denoting" (1905). The apparent or grammatical form of, for example, *The present King of France is bald* is subject–predicate, but Russell argued that its logical form is very different, expressible as: *Something is a King of France, not more than one thing is a King of France, and whatever is a King of France is bald*. In Russell's opinion, being right about logical form has instrumental as well as intrinsic value: it prevents unsatisfactory metaphysics, for example a Meinongian ontology driven by a supposed need to postulate a being corresponding to "the King of France", and it makes possible plausible epistemology, since one can know that there are things with certain properties without being acquainted with any examples. In Russell's opinion, philosophical logic was not some specially important area of philosophy but the core of all serious philosophising:

> Every philosophical problem, when it is subjected to the necessary analysis and purification, is found either not to be really philosophical at all, or else to be, in the sense in which we are using the word, logical. (Russell 1914: 42)

Later in the essay, the sense in question is identified precisely as "philosophical logic".

Russell applied the philosophical logic programme to a wide range of problems, most centrally in the philosophy of mathematics. His Logicism is the claim that mathematical truths are logical truths, provable by logical means (if provable at all), and so knowable in the same way as logic. The core of the position can be represented as a claim about logical form: we get closer to the logical form of, say, *2+2=4* by seeing it as about relations between classes of classes, and closer still when we appreciate that the apparent reference to classes is inessential (the "no class" theory of classes). This leads to his draconian view that there are no mathematical objects, neither numbers nor classes, or rather, more cautiously and correctly, to the view that there is no reason to suppose that there are any. In the Russellian tradition, logical form proposals are never far removed from often dramatic metaphysical or epistemological claims.

Other applications of the programme of philosophical logic were to belief and judgement, truth, existence (including the ontological argument for the existence of God), and the epistemologically motivated programme of replacing "inferred entities" by "logical constructions", applied to the external world and to minds.

Russell's influence is at work in any philosopher who is willing to make a contrast between grammatical form and logical form. We find a clear statement, and an acknowledgement to Russell, in Wittgenstein's Tractatus:

> 4.002: Language disguises thought. So much so, that from the outward form of the clothing it is

impossible to infer the form of the thought beneath it, because the outward form of the clothing is not designed to reveal the form of the body, but for entirely different purposes.

4.0031 It was Russell who performed the service of showing that the apparent logical form of a proposition need not be its real one.

Russell's influence is also at work in many of those who pepper their text with logical symbols purportedly in the service of paraphrasing some idiom of ordinary speech. These philosophers may not be explicitly soldiering under the banner of the philosophical logical programme, but they reveal that they have absorbed its influence, perhaps barely consciously, if they hold that the symbolic paraphrases reveal something that would otherwise be concealed in what they paraphrase. The upshot is that a fully detailed history of philosophical logic in this century would embrace a very large part of analytic philosophy. Rather than attempt this, I will in §2 indicate some major events in the development of philosophical logic, and will consider its methodology, and the justification for the conception of logical form which philosophical logic requires. In §3 I will discuss an aspect of the concept of truth which belongs to philosophical logic: the paradoxes to which it gives rise, and how these are related to some other paradoxes. I conclude with a prognosis concerning the future of philosophical logic.

2.   "Logical form": the holy grail of philosophical logic

2. 1. The Russell–Quine–Davidson tradition

Russell thought that the language of *Principia Mathematica* was the language of logical forms: in this language, there is no distinction between logical form and grammatical form, no ambiguity and no lack of clarity. A major philosopher to inherit this view, fully explicitly, was W. O. Quine, whose "On what there is" (1948) was a milestone in the application of logical form to ontology. The main claim of that paper was that a theory's ontological commitments are to be judged by first regimenting it into the language of first order logic (a proper part of the language of *Principia Mathematica*) and then seeing what must exist for the existential quantifications to be true: "To be assumed as an entity is, purely and simply, to be reckoned as the value of a variable" (Quine 1948:13). The regimentation is essential to prevent us, for example, from supposing that commitment to the view that a person acted for another's sake is thereby commitment to an ontology of sakes: such false appearances would disappear once the claim is regimented in first order logic.

Although a distinction between grammatical and logical form was freely used in the analytic tradition, for many decades the methodology of the distinction was relatively little discussed. In "On referring" (1950a), Peter Strawson entered some doubts, deriving from his sense of the plasticity and context-dependence of ordinary language compared to the rigidity of formal languages. He suggested that there was no adequate programme of finding logical forms, for "ordinary language has no exact logic" (1950a: 27). He was vividly aware that the same words can be used in different contexts to say different things ("I am hungry", said by me, says that I am hungry, but said by you it says that you are), and this precludes an exact description of language that pays no attention to how it is used. By contrast, formal languages are generally devised in such a way that there is no context-dependence; what might in some ways resemble context dependence will be brought under the thumb of model-theoretic precision, for example, the fact that a variable can be assigned different objects on different interpretations.

The fullest defence of the Russell–Quine approach to logical form is due to Donald Davidson. He made free use of the contrast between logical and grammatical form in "The logical form of action sentences" (1967), proposing that sentences like "Shem kicked Shaun" are "really" (that is, at the level of logical form) existential quantifications over events (kicks), and that "kicked" is, despite being apparently two-place, in reality three-place. The logical form is something like: there is a kick which Shem kicked to Shaun. James Cargile (1970: 137–8) protested that, look as he might, he could not see the existential quantifier, and could not imagine a basis for supposing that "kicked", as it occurs in the sentence in question, has more than two places. More generally, he stressed that a claim of logical form should amount to more than a claim of logical equivalence, even if one of the equivalents is in some sense more perspicuous than the other. This elicited from Davidson historically the first full account of criteria for selecting logical forms (Davidson 1970), and a justification of their interest. At the same time, Davidson was explicit that finding logical forms is not, as Russell thought, the end of the story, but only the beginning: after that, and only after that, the substantive job of philosophical analysis can begin.

Davidson's views have been developed (for example, by Lycan 1984), but in the purely philosophical tradition there is no radically distinct paradigm in sight. A notion of logical form has led a somewhat different life in the hands of linguists, where it is often referred to as a level of syntactic description called "LF". As early as 1957, Noam Chomsky suggested that a proper description or representation of a sentence, from the grammatical point of view, would have many layers, and this paved the way for the idea, developed subsequently, that LF would be one such layer. A standard view is that LF is the level of linguistic representation at which all grammatical structure relevant to semantic interpretation is specified (cf. Hornstein 1995). Though there is no doubt that the motivations for introducing LF typically differ from the motivations which led philosophers to introduce logical form, the question whether the LF of linguists and the logical form of philosophers are or are not essentially the same under the skin is a vexed one, which we will not explore (cf. Neale 1993). Instead, I will examine in more detail some of the methodological issues raised by the conception of logical form that has been dominant in the philosophical as opposed to the linguistic literature (though the distinction is by no means hermetic).

2.2 .Russell

We know what Russell's kind of philosophical logic is if we know what logical form is, for philosophical logic is the programme of finding logical forms. Here is Russell's earliest explicit formulation:

> Form is not another constituent [of propositions], but is the way the constituents are put together.
> It is forms, in this sense, that are the proper object of philosophical logic. (Russell 1914: 52)

The basic form is the atomic form, exemplified by a sentence composed of an *n*-place predicate conjoined with *n* names. Though this may now strike us as scarcely worthy of comment, this conception of the basic form was innovative and controversial. As Russell made explicit, the notion of an atomic sentence is designed to supersede the older notion of subject–predicate form, which Russell believed was defective for two reasons. One explicit reason is that it does not do justice to relations. Treated as subject–predicate, "Mary loves John" and "John loves Mary" share neither a subject nor a predicate (for "loves John" is a different predicate from "loves Mary"). Regarded as a two place relation, there is a relation in common and two common names, and the only difference is order. The second description is superior from the point of view of developing a logic of relations.

The other reason for dissatisfaction with the notion of subject–predicate as opposed to atomic sentences was implicit. In the kind of languages with which Russell was concerned, there is a finite stock of names and predicates and hence a finite stock of atomic sentences. In the traditional conception, subject–predicate sentences are in infinite supply, for subject expressions can be complex without upper limit. For example, definite descriptions like "The Queen of England" are counted as subject expressions, which means that one must also count as subject expressions the following "The Queen of the country which colonized India", "The Queen of the country which colonized the country in which alone tigers are indigenous", and so on. It was important to Russell that language or thought should start with a finite number of elements, and this is ensured by the conception of an atomic sentence or atomic judgement; the conception of a subject–predicate judgement cannot play the same role.

Russell's enthusiasm for Frege's insistence upon the notion of the atomic sentence[1] was in part based on his appreciation of the way in which it enabled the provision of the truth conditions of quantified sentences in terms of quantifications over the truth conditions of atomic sentences. As Dummett describes this aspect of Frege's work:

> "everybody envies somebody" is true just in case each of the sentences "Peter envies somebody",
> "James envies somebody", … is true; and "Peter envies somebody" is, in turn, true just in case at
> least one of the sentences "Peter envies John", "Peter envies James", is true. (Dummett 1973a: 11)

We find an account of quantification along these lines in Russell's "On denoting", as a precursor to his classification of definite descriptions as quantifier phrases. This contrasts with the traditional approach, found for example in Leibniz, according to which the inferential role of sentences like "Socrates is mortal"

---

[1] "Traditional logic regarded the two propositions 'Socrates is mortal' and 'All men are mortal' as being of the same form; Peano and Frege showed that they are utterly different in form. … the philosophical importance of the advance which they made is impossible to exaggerate" (Russell 1914: 50).

is explained in part by their equivalence to "Some-or-every Socrates is mortal" (cf. Sommers: 1982: 15–21).

The constituents of an atomic proposition or judgement are (i) the bearers of the names and (ii) the property or relation introduced by the predicate; the logical form is not a constituent, but is, rather, the connection which holds between the constituents in virtue of which the sentence manages to say something. Russell struggled with this fundamental notion, labelling the problem that of the "unity of the proposition". Take an atomic sentence like "Desdemona loves Cassio", and suppose that it belongs to the world of fact rather than fiction.[2] The three constituents are Desdemona, Cassio and love, three real entities (a woman, a man and a relation). But the sentence is more than just a list of these things. It "combines" the things in such a way that the sentence says (truly or falsely, and in fact falsely) that Desdemona loves Cassio; in this it achieves something which no mere list can achieve. Russell chose an example of a falsehood to bring home the difficulty: in the case of a truth, this process of combination could be held to result in a fact, a complex entity consisting of Desdemona's love for Cassio. But as the sentence is false, there is no such fact, no such complex entity. The problem was to understand how the combining of the constituents could result in an entity which made the sentence more than a list without resulting in a fact, and so, absurdly, precluding the possibility of falsehood. At various points, Russell appealed to logical form to answer this question, though he never arrived at a formulation which satisfied him (cf. Russell 1913). The constituents of, say, a one-place atom are welded together through the unary atomic form. This form could not itself be a constituent on pain of regress (how is the form bound to the other constituents?); but it seemed puzzling to Russell that it could do any work in the proposition without being a constituent. And indeed it seemed puzzling to him how anything could combine the constituents in a way which made them say something without also combining them in a way which made what they said true, and so precluded falsehood. As he puts it, we need the relation to "really relate" the terms if we are to have more than a list; but if it really relates the terms then the proposition is true.

Russell's difficulty is not much discussed nowadays outside Russell scholarship, but it seems to me that it touches many contemporary theories. To take just one example: some identify propositions with sets of possible worlds, and say that a proposition is true at a world just in case that world belongs to it. This leaves unanswered the question of how a set can say something. Not all sets do so; only those which are propositions. What additional features do such sets possess? I do not say that the question is unanswerable, but only that it remains a genuine question.[3]

2.3. Three strands in the conception of logical form

Russell's conception of logical form thus had a shaky foundation in the account of atomic sentences and propositions. It was supposed to constitute the cement that bound the constituents of a proposition together in such a way that they said something, formed a content that could be judged, but Russell never felt he had satisfactorily explained how this could be so. Despite this, the remainder of the account proceeded fairly smoothly, thanks to a shift of focus. The standard sentence connectives bind atoms together in a way that can be explained through their truth functionality: they are pure cement and do not themselves introduce further constituents. The early Wittgenstein advertised this as a central tenet of the *Tractatus*:

> 4.0312 My fundamental idea is that the "logical constants" are not representatives; that there can be no representatives of the logic of facts.

Logical constants, like "and" and "if" (or "&" and "→"), Wittgenstein says, are not representatives, that is, they do not stand for anything, do not introduce a further constituent but play a purely cementing role.[4]

---

[2] We must also pretend that "Desdemona" and "Cassio" are proper names in the logical sense, that is, are not to be analysed as really definite descriptions.

[3] The way in which I attempt to answer it (2002: V) on Russell's behalf could be adapted to a range of views about the nature of propositions.

[4] Wittgenstein writes as if he thought that Russell disagreed and regarded logical constants as non-linguistic entities: "it becomes manifest that there are no 'logical objects' or 'logical constants' (in Frege's and Russell's sense)" (1921: §5.4).

Given that the role of the truth functional connectives is explained in terms of the truth functional dependence of the truth values of complexes upon that of their sentential parts, why should not Russell, or the early Wittgenstein, have regarded them as referring to truth functions? On the face of it, this would have shown how a constituent could also function as cement, for a function binds its arguments together. Though I am not aware of an explicit discussion in Russell, I think his motivation was probably ontological. Consider the fact that Plato was Greek and Napoleon was French. How must the world be for this to obtain? It is necessary and sufficient that two simpler facts exist: that Plato was Greek, that Napoleon was French. On the face of it, there is no need for a third thing to exist, namely the conjunction truth function. So there is no ground for supposing that there is such a function, and we do better to explain the truth or falsehood of conjunctive sentences without appeal to these doubtful entities.

Logical form as cement: that is one strand in its nature. It goes with the idea that logical vocabulary or the logical constants serve to introduce just cement, and not constituents, and we will see that this is one strand of which much is made in more recent writing, especially by Davidson. Another strand is logical form as revelatory of ontological commitment, the strand which Quine most famously exploited. The third strand is logical form as revelatory of the features of a judgement most relevant to inference. For example, Russell thought that Plato's problem of non-being will simply not arise if the problematic sentences are accorded their proper logical form, for then the paradoxical inferences are unavailable. The problem is to explain how can one truly say that something does not exist, given that if one says this of something one says something false, whereas if one says it of nothing one says nothing at all. The first horn of the dilemma presupposes that the logical form of a statement of non-existence is "not (Exists (a))". This does indeed classically entail that a exists (i.e. that there is something identical to a). According to Russell, a statement of non-existence does not have this form, but rather something along the lines: "not (there are so-and-sos)". This neither supports the unwanted inference, nor can it be regarded, as the second horn suggests, as "saying nothing". So "Unicorns don't exist" has the unproblematic logical form "not (there are unicorns)", and " Vulcan does not exist" has the unproblematic logical form "not (there is a unique planet between Mercury and the sun which is responsible for the perturbations in Mercury's orbit)".[5]

Can the three strands we have identified in logical form be seen as part of a single phenomenon? Cement provides structure, binding the bricks together, and valid inference, or at least logically valid inference, is inference in virtue of form or structure, independent of the specific features of the constituents. "Fido barks, therefore something barks" is valid in virtue of its form, which is shared by countless other arguments ("Plato thinks, therefore something thinks", "Napoleon commands, therefore something commands", and so on). So the logical vocabulary both binds and also shows the features relevant to inference; this connects two of our three strands. Since existential generalization is a mechanical matter once the premises are presented in their correct logical form, there is evidently a close connection with ontology, thus securing our third strand. Whereas we might think we could apply existential generalization to "The present King of France does not exist" to yield the undesirable "Something does not exist", no such inference is available once the true form of the premise is appreciated ("Nothing is uniquely king of France").

2.4. Davidson

This was how logical form was considered for about half a century from before the first world war. Donald Davidson redeveloped the notion, first in his philosophical practice (Davidson 1967: "The logical form of action sentences"), and shortly thereafter in an explicit statement of the methodology (Davidson 1970: "Action and reaction"). He begins with what I regard as a fully accurate statement of the issue:

> … logical form was invented to contrast with something else that is held to be apparent but mere: the form we are led to assign to sentences by superficial analogy or traditional grammar. What meets the eye or ear in language has the charm, complexity, convenience and deceit of other conventions of the market place, but underlying it is the solid currency of a plainer, duller structure, without wit but also without pretence. This true coin, the deep structure, need never feature directly in the transactions of real life. As long as we know how to redeem our paper we

---

[5] The claim that proper names like "Vulcan" have as their logical forms definite descriptions like "the planet between Mercury and the sun which is responsible for the perturbations in Mercury's orbit" has been much disputed within the philosophical logic tradition: see Stanley's essay in this volume. [Ed.: to check]

can enjoy the benefits of credit. … The image may help to explain why the distinction between logical form and surface grammar can flourish without anyone ever quite explaining it. (Davidson 1970: 137)

In considering how the distinction could be justified, Davidson begins by locating the connection between logical form and inference: "Part of the answer—the part with which we are most familiar—is that inference is simplified and mechanized when we rewrite sentences in some standardized notation" (Davidson 1970: 138). But the desire to simplify and mechanize inference does not in itself underwrite the contrast between logical and merely surface form. If simplifying inference was all that was at stake, one could as well say: "let's recast our thoughts in this simple language, good for revealing inferential relations", without pretending that the recasting provided a better account of the shape of the original moulds.

Davidson's final position is that the logical form of a sentence in natural language is the result of doing to it whatever is necessary to produce something accessible to systematic semantic theory:

To give the logical form of a sentence is, then, for me, to describe it in terms that bring it within the scope of a semantic theory (Davidson 1970: 144).

This redescription may involve no change: perhaps "Fido barks" is the logical form of this sentence. Or the redescription may involve just "one small point". According to Davidson's paratactic account of the logical form of sentences ascribing propositional attitudes, the logical form of "Plato said that Hesperus is visible in the evening" is "Plato said that. Hesperus is visible in the evening". The only difference is the full stop (the small point) after "that", this last being construed as a demonstrative pronoun referring forward to the utterance of "Hesperus is visible in the evening". Or the change may be considerable. In the early paper (Davidson 1967) he suggested that the logical form of "Shem kicked Shaun" is something like "There was a kicking, x of Shaun by Shem"; in (roughly) first-order notation:

$\exists x(\text{kick }(x) \text{ \& kicked (Shem, } x, \text{ Shaun)}).$

Ultimately, the justification for a logical form proposal, within Davidson's approach, is that it presents the original sentence, or its content, in a way which makes it accessible to systematic semantic theory. On the assumption that such a theory will be cast somewhat in the style of a first order truth theory, and so will be extensional, the seemingly non-extensional idioms of belief-ascription are not directly accessible to semantic theory. The paratactic account is a dazzling solution to this problem: the seeming non-extensionality is an illusion of grammatical form, a result merely of the missing point after "that". A sentence is non-extensional iff substitution of co-extensive expressions within it does not preserve truth value. If we think of the original belief attribution as a single sentence, "Plato said that Hesperus is visible in the evening", it is possible that substituting the co-extensive "Phosphorus" for "Hesperus" would alter the truth value. It might be that "Plato said that Hesperus is visible in the evening" is true, but "Plato said that Phosphorus is visible in the evening" is false. Once we see that there are really two sentences, the illusion of non-extensionality is dispersed. In the first sentence, "Plato said that", "that" will refer to different entities according to what utterance immediately follows; and replacing "Hesperus" by "Phosphorus" will not change the truth value of the second sentence ("Hesperus is visible in the evening"). There is no shift of truth-value, within any one sentence, so long as there is no shift in the extension of any of that sentence's parts.

Accessibility to semantic theory is the ultimate justification. For Davidson, this is intimately connected with inference, as this way of putting the point illustrates:

By saying exactly what the role [of a verb] is, and what the roles of the other significant features of the sentences are, we will have a deep explanation of why one sentence entails the other, an explanation that draws on a systematic account of how the meaning of each sentence is a function of its structure (Davidson 1970: 142).

For example, appreciating the extra argument place in the logical form of sentences containing verbs of action permits an understanding of inferential relations in antecedently familiar terms. Intuitively, Davidson says, the following inference is not merely valid, but valid in virtue of form:

John buttered the toast in the bathroom, therefore John buttered the toast.

On many views of the logical form of the premise, the validity of this inference has to be accepted as a brute and inexplicable fact. Perhaps there are two verbs "buttered", one 3-place (and used in the premise) and one 2-place (and used in the conclusion) and a "meaning postulate" tells us that we can always infer

from a sentence dominated by the 3-place predicate to a corresponding sentence dominated by the 2-place one. Apart from the fact that such an approach would have to postulate an indefinite series of predicates "buttered" (a 4-place one for "John buttered the toast in the bathroom with a knife", a 5-place one for "John buttered the toast in the bathroom with a knife at midnight", and so on), it is also explanatorily weak compared to Davidson's proposal, which is that the argument has the form:

$$\exists x(\text{Buttered(John, the toast, } x) \ \& \ \text{In(the bathroom, } x)) \vdash \exists x(\text{Buttered(John, the toast, } x))$$

The validity of the argument is subsumed under a familiar first order case: dropping a conjunct under existential quantification; this provides a serious explanation of its validity.

2.5. Russell and Davidson compared

Russell would no doubt have been pleased with this result. But Davidson's view differs significantly from Russell's. First, Davidson's concern is with the logical form of sentences, whereas Russell's is with the logical form of facts, judgements, or propositions (regarded as non-linguistic things). This connects with the second difference, that whereas for Russell the primary constraint on logical form is imposed by the need to do justice to inferential relations between judgements or propositions, for Davidson the primary constraint is imposed by the need to do justice to the semantics of sentences. Thirdly, whereas Russell claimed (perhaps exaggerating his real view) that philosophical logic was the whole of philosophy, Davidson said that finding logical forms was just the beginning: after that there remained the serious business of philosophical analysis.

Davidson aimed to derive the relation between logical form and logical inference from the way in which semantic obligations determine logical form. A semantic theory is a theory of truth, with theorems like "'Snow is white' is true iff snow is white", which Davidson calls "T-sentences". One sentence entails another if the truth of the first guarantees the truth of the second, and this is formal (a matter of form) if this is so in virtue of the structure of the sentences. Semantic theory will reveal this structure and so underwrite formal entailments.

The connection between truth theoretical semantics and logical entailment emerges in Davidson's account of the logical constants. He said that these are expressions which receive a recursive rather than a basis clause within a semantic theory. "And", for example, will have a semantic axiom like the following:

for all sentences, p, q, 'p and q' is true iff p is true and q is true.

This is recursive in the sense that it specifies a condition for the truth of a conjunction in terms of the truth of its components, so it is only one step on the way to the goal of a specification of a truth condition in truth-independent terms. If we want to reach a "pure" truth condition for, e.g. "Plato thinks and Napoleon commands" we may first apply the axiom for "and" to get:

"Plato thinks and Napoleon commands" is true iff "Plato thinks" is true and "Napoleon commands" is true

but we will also have to find truth conditions for the components "Plato thinks" and "Napoleon commands" so that we can use the equivalences

"Plato thinks" is true iff Plato thinks
"Napoleon commands" is true iff Napoleon commands

to derive the target T-sentence

"Plato thinks and Napoleon commands" is true iff Plato thinks and Napoleon commands.

By contrast a semantic axiom for an expression like "Plato" or "thinks" will be a "basis" clause, one which does not involve the notion of truth on the right hand side, for example:

"Plato" refers to something iff that thing is Plato
"thinks" is true of something iff that thing thinks.

The contrast between the logical constants (with their recursive clauses) and other expressions (with their basis clauses) mirrors at the linguistic level Russell's contrast between cement (logical constants) and bricks (non-logical expressions). The contrast will be written into logical forms. To the extent that the validity of inference is grounded entirely in structure and independently of the bricks that compose it, identifying logical forms will be the same as identifying inferential properties. Hence Davidson's use of facts about inference to test for the correctness of logical form proposals:

> … my only reasons for "rendering" or "paraphrasing" event sentences into quantificational form was as a way of giving the truth conditions for those sentences within a going theory of truth. We have a clear semantics for first-order quantificational languages, and so if we can see how to paraphrase sentences in a natural language into quantificational form, we see how to extend a theory of truth to those sentences. Since the entailments that depend on quantificational form can be completely formalized, it is an easy test of our success in capturing logical form within a theory of truth to see whether our paraphrases articulate the entailments we independently recognize as due to form. (Davidson 1970: 144)

Whereas for Russell, the connection between logical form and inference was basic and primitive, Davidson seeks to derive it via the connections between logical form and semantics, and between semantics and entailment.

A further difference between Russell's and Davidson's perspective on logical form relates to identity. For Russell, whose logic involved quantification over predicates of every type and order, "=" is introduced as a defined logical connective:

*13.01 $x = y =_{df} \forall\varphi(\varphi!x \leftrightarrow \varphi!y)$. [6]

This is not available within Davidson's predominantly first order perspective. "=" will receive a basis clause and not a recursive clause in semantic theory and so should not be counted as a logical constant. This has further consequences. Whereas Russell can regard a purely numerical sentence like "There are exactly two entities" as composed wholly of logical vocabulary (a formalization is: $\exists x \exists y[x \neq y \ \& \ \forall z(x=z \lor y=z)])$, Davidson cannot. This precludes Davidson from thinking of standard numerical quantifiers (of the form "there are exactly n so-and-sos") as belonging to logic.

2.6. Inference and logical form

We have seen that Russell thought that logical form was the form of a judgement, not of a sentence. This explains his subsequent view of his work in the period up to the end of the first world war as work that ignored language (Russell 1959: 145). This opinion surprises those who anachronistically see the early Russell, for example in his Theory of Descriptions, as engaged in an account of the semantics of definite descriptions in natural language. Even if we agree that the Theory of Descriptions, regarded as an account of natural language, involves "the butchering of surface structure" (Evans 1982: 57), this is no criticism of Russell, who was not concerned with surface structure but only with giving the most revealing account of judgement, of what is before the mind of a thinker using a sentence containing a definite description.

Up to now, I have drawn a veil over how precisely Davidson understands the relation between a sentence of natural language, s, and its logical form, s_{LF}. There are two options. He can say that s has as its logical form s_{LF} iff the T-sentence for s simply is "'s' is true iff s_{LF}". Or he can say that s has as its logical form s_{LF} iff as a preliminary to applying a semantic theory to s it is translated as s_{LF} and then the T-sentence is "'s_{LF}' is true iff s_{LF}" (the T-sentence is, as Davidson calls it, "homophonic"). The first option, as Davidson suggests in the following quotation, makes good sense of such claims as that an action sentence "really" contains an existential quantifier:

> all I mean by saying that "Jones buttered the toast" has the logical form of an existentially quantified sentence, and that "buttered" is a three-piece [sic] predicate, is that a theory of truth meeting Tarski's criteria would entail that this sentence is true if and only if there exists ... etc. (Davidson 1970: 143)

However, this first option is precluded by the methodology. If there is to be any chance of deriving a T-sentence for "Jones buttered the toast" whose right hand side is as Davidson envisages, then the axiom for "buttered" must treat this as a 3-place predicate, in which case the T-sentence must refer, on its left hand side, not to the unchanged English sentence but to its "logical form": "there exists an x such that (Buttered (Jones, the toast, x))". So Davidson has to adopt the second option, on which it is the natural language sentence's logical form, which will often be a sentence rather different from the natural language sentence itself, that is addressed by the semantic theory. This raises the question of how the semantics for s_{LF}, which we can accept are impeccably specified by the truth theory, transfer to the distinct sentence s. The English

---

[6] The "!" serves to restrict the quantifier to "predicative functions", ones which do not infringe the Vicious Circle Principle (see below).

sentence and its logical form may agree in their truth conditions, yet achieve them through different semantic mechanisms. Cargile's question appears still to require an answer.

It would be appropriate for Davidson to reaffirm the role of inference. The original sentence and its logical form agree on all "formal" entailments (as well as on truth conditions), and this suggests that they stand in an intimate relationship, say R. R needs to be so intimate that, if it holds between two sentences, a semantic story told for the one is also correct for the other (even though just one of the sentences can be the target of a T-sentence). In my opinion, it remains for the Davidsonian methodology of logical form to specify an R which both justifies the claim that if $R(s_1, s_2)$ then correct semantics for s_2 are thereby correct semantics for s_1, and also relates "John buttered the toast" and "there exists an x such that (Buttered (Jones, the toast, x))".

As I said above, Davidson is explicit that logical form is at most the beginning of philosophy. Once we have found the logical form, we can go on to do the philosophical analysis and metaphysics. In the case of action sentences, logical form considerations tell us that they quantify over events, but are silent on the nature of events: their criteria of existence and identity. In the case of sentences ascribing propositional attitudes, logical form considerations tell us that these involve a relation between a person and an utterance, but they do not tell us what this relation is. Davidson suggests that if I truly utter a sentence like "Plato said that Hesperus is visible in the evening", then Plato said something which makes him and me samesayers in virtue of my utterance of "Hesperus is visible in the evening". This claim belongs to analysis, not to logical form, and as Davidson says, "samesaying" itself marks a place where further analysis is needed (though he is optimistic that his conception of semantic theory can play a useful role).

In the course of a century, the conception of philosophical logic has changed from a panacea for philosophical problems to an aspect of semantics. In addition, a new usage has sprung up, according to which philosophical logic is the investigation of logical concepts and problems that are of special relevance to philosophers, truth being a prominent example (as we will see in the next section). Some insist on marking the divergent streams with distinct terms: keep "philosophical logic" for the logical form programme, understood on Russellian or Davidsonian lines, and "philosophy of logic" for philosophical discussion of the notions special or central to logic, like truth, entailment, and form. There is a natural tendency to blur the boundaries, for philosophical logic (in the first sense) makes essential use of the notions of truth, entailment and form, as we have seen, and it is in the nature of philosophy constantly to question the very tools it uses. Among these tools, nothing would seem more important than the notion of truth; and in the twentieth century, this has been a rich source of puzzlement, paradox, and imaginative and challenging theorizing.

3. Truth

3.1   Some approaches to truth

The concept of truth has played a central role in most philosophical traditions, and very conspicuously so in analytic philosophy in the twentieth century. Whatever their central interests, philosophers cannot ignore truth, and it has played an especially important part within philosophical logic on account of the paradoxes with which it is associated, and the theoretical work arising from reflection on these paradoxes. Whatever else one may demand of logic, consistency would seem to be an overriding requirement; yet the notion of truth has appeared to lead swiftly to contradictions. These are the topic of subsequent subsections (§3.2–§3.5).

One view of truth, which has informed many and disparate theories, goes back to Aristotle:

> To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true. (*Metaphysics* Γ7, 27)

One can classify a variety of opinions about truth in terms of how they react to this Aristotelian view. Among those who agree with Aristotle, there are those who think that it shows that the nature of truth consists in some kind of correspondence with reality, and those who think that it shows that truth has no

"nature" but is some kind of eliminable or at most very "thin" notion, expressing no substantive property. Those who disagree with Aristotle think that an account of truth should be sought in quite different terms. Non-Aristotelian notions which have been used in explicating truth include coherence, convergence and assertibility. Some have held that to be true is to be consistent with what we believe; or that the truth is that upon which best opinion will in the long run converge; or that what is true coincides with what, under the most favourable conditions, we would be justified in asserting. Important as these approaches are, the following remarks relate just to the Aristotelian conception.

Some philosophers see Aristotle as advancing a correspondence theory of truth: what is true is what corresponds to the facts. One cannot dispute that in ordinary parlance there is little to choose between calling something true and saying that it corresponds to the facts. This as such does not give correspondence any special role to play in explaining truth: one could as well say that the notion of correspondence is to be explained in terms of truth. For a correspondence theory to make a significant contribution, it needs to offer more than just the verbal equivalence noted; it needs to show that the notions of correspondence and fact are somehow more basic than that of truth, and one way to achieve this is to give an independent account of the relation of correspondence itself, and of the terms between which the relation is supposed to hold, the truth-bearer on the one side and the truth maker on the other. Sentences or beliefs are standard candidates for truth bearers, and facts, states of affairs, or simply reality are candidates for truth makers.

Russell (1912) suggested that truth bearers are beliefs, analysed as relations holding between believers and various things in the world. If Othello believes that Desdemona loves Cassio, the belief consists in a relation between Othello, Desdemona, love, and Cassio. This belief is true if there is a fact whose constituents are Desdemona, love and Cassio (in that order). This provides an atomistic version of the correspondence theory. Truth consists not in correspondence with reality construed in an undifferentiated way, for example, simply as the way things are, but in a kind of isomorphism between parts of the belief and specific parts of reality.

Russell himself was never satisfied with this account, on the grounds that it did not do proper justice to love. In Othello's belief, this features as a term in a relation; but if there is a corresponding fact, love needs to feature in a quite different way, as a binder together of terms rather than as itself a term. Here we encounter an aspect of the problem, already noted above, of the "unity of the proposition".

The problem of giving an adequate characterization of the truth maker was at the centre of a famous debate between John Austin and Peter Strawson. Strawson claimed, reasonably enough, that a serious correspondence theory should be able to give a general specification of truth makers without invoking the concept of truth, and on which they were clearly independent of truth bearers. Austin's truth makers were facts, but Strawson claimed that these were really linguistic entities, and not suitably distinct from truth bearers (which both parties referred to as statements): "facts are what statements (when true) state.… If you prize the statements off the world you prize the facts off it too; but the world would be none the poorer" (1950b, p.38-9).

In developing the details of his theory of truth, Austin (1950) gave a quite theoretical account of how the correspondence was supposed to work.[7] He argued that a sentence is linked to two kinds of thing in the world, by two different sets of conventions. The descriptive conventions link it to a state of affairs, something which may or may not obtain. The demonstrative conventions link it to a particular situation. Thus (to use Austin's example) an utterance of the sentence "The cat is on the mat" is linked by the descriptive conventions to the state of affairs of the cat being on the mat, and by the demonstrative conventions to some particular situation in some region of the world. (Austin is rather unclear about just what particular situation is involved, as Strawson stresses.) The utterance of the sentence is true just if the particular situation is of a kind with (that is, instantiates) the state of affairs. For many years after this theory was advanced, it was widely thought to have been refuted by Strawson. (Strawson objected that Austin had made truth purely conventional, but this seems to have been a misunderstanding.) However,

---

[7] Perhaps it was unfair of Strawson to claim that Austin had given no truth-free specification of the nature of facts: see Austin (1961).

Austin's theory has received a new lease of life in the hands of Barwise and Etchemendy (1987), who claim that it can provide an account of truth for which the Liar paradox does not arise (this paradox is discussed below, starting in §3.2).

One of the most famous and important approaches to truth within the Aristotelian framework is Alfred Tarski's semantic conception of truth, of which he says that it certainly conforms with Aristotle's conception (Tarski 1944: 72). In an attempt to make Aristotle's definition more precise, he gives an example of how it might apply to the sentence "snow is white", suggesting that "the sentence is true if snow is white, and … false if snow is not white", in summary:

>   The sentence "snow is white" is true if, and only if, snow is white. (1944: 63)

Tarski goes on to suggest that a correct definition of truth will entail every equivalence of this kind. The general form of the crucial equivalences is given by:

>   T   X is true if, and only if, p

where "X" marks the place for some expression which refers to a sentence in a way which makes manifest how the sentence is built up of words,[8] and "p" marks a position occupied by that very sentence (the one referred to by the expression in "X"-position).[9]

>   we wish to use the term "true" in such a way that all equivalences of the form (T) can be asserted, and *we shall call a definition of* truth *"adequate" if all these equivalences follow from it*. (Tarski 1944: 63)

For a formalized language, he was able to provide a precise definition of truth which is adequate by this standard, set in a forbiddingly technical theory (Tarski 1933). No one disputes his success in this respect, but there is widespread dispute about the significance of the result. I mention four debated points.

(1) Why is not (T) itself already a definition of truth? Tarski's answer was that it is not a definition because it is not even a sentence, but only the schema of a sentence. "X" has no specific reference, and there is no sentence that we are supposed to substitute for "p". Obvious attempts to turn (T) into a sentence lead to nonsense. For example, "for every sentence X, and every sentence p, X is true iff p" is strictly nonsense, because quantifiers, as Tarski conceived them, are associated with individual variables, expressions which occupy the position of names, but the position marked by "p" is not such a position (it is a position to be filled by a sentence, not by a name). Even if we could understand this attempted generalization, it seems to point in an absurd direction, suggesting that for any pair of sentences, one gives a necessary and sufficient condition for the truth of the other.

(2) The T-equivalences seem trivial, so how could it be enough for a definition of truth to be correct that it entail them all? On Tarski's side, it must be said that his criterion of correctness ensures that "true" will apply to all and only true sentence; this is so whatever we think about the triviality or significance of T-equivalences. On the other hand, one might want more from a definition, or at any rate from a philosophical account, than that it get the extension of "true" right. One might, not unreasonably, want an informal account of how truth relates to other interesting philosophical notions like reality, objectivity, warrant and assertion, and an explication of why truth is of value and why we should strive for it. I imagine that Tarski would accept this observation, while insisting that these other tasks can be best performed only once one has provided a correct definition, in his narrower sense of something which precisely determines the extension of "true".

---

[8] There is no need for a correct definition of truth to entail, for example, that the first sentence of the third section of Sainsbury's essay on philosophical logic is true iff the concept of truth has played a central role in most philosophical traditions.

[9] This is a narrower condition than the one earlier used by Tarski (1933), which required only that the sentence in the position of "p" translate the sentence named by the expression in "X"-position. The earlier position is more liberal in that identity is presumably not the only form of correct translation.

(3) The T-equivalences relate sentences to reality (for example, to snow's being or not being white) and so the semantic conception "finds itself involved in a most uncritical realism" (Tarski 1944: 75). Thus put, the objection is silly, and Tarski deals with it briskly: the semantic conception has nothing to say about what it is for snow to be white, and so a fortiori says nothing that is specially realist, irrealist, empiricist or anything else:

> we may accept the semantic conception of truth without giving up any epistemological attitude we may have had; we may remain naïve realists, critical realists or idealists, empiricists or metaphysicians—whatever we were before. The semantic conception is completely neutral toward all these issues. (Tarski 1944: 75)

However, the objection can be urged from a different quarter. Michael Dummett has suggested that to accept the Law of Excluded Middle for a given subject matter (the Law that for every sentence, either it or its negation is true) is to adopt realism with respect to that subject matter. For one is then committed to the view that even if we have no conceivable means of finding out which if either of a sentence and its negation is true, nonetheless one of them is: reality may go beyond what we can conceivably know. (See for example Dummett 1973b, 1978: Preface.)

(4) Don't the T-equivalences show that truth is redundant? Instead of saying "'Snow is white' is true" we can just say "Snow is white"; we can do just as well without the predicate "true". Tarski responds by giving two kinds of case which cause trouble for this view. One is where we ascribe truth without knowing the content of that to which we ascribe it, for example:

> The first sentence written by Plato is true. (Tarski 1944: 73)

Since we do not know what Plato's first written sentence was, we cannot affirm it instead; but the displayed claim seems intelligible and, indeed, mildly interesting. The other problematic kind of case involves generalizations like

> All consequences of true sentences are true. (Tarski 1944: 73)

The task of specifying every case in which the relation of consequence obtains is too long to be completed, whereas the displayed sentence says something important. It might seem we could get rid of "true" in such cases by a device along these lines:

> for all s, p, if s, and p is a consequence of s, then p.

If these quantifiers ("for all s, p") are of the kind Tarski recognized, this is nonsense for a reason already noted: the variables should mark a position for a name, whereas in the above the relevant position is one to be filled by a sentence. There may be other kinds of quantification. Perhaps one could introduce a sentence-quantifier, whose variables are to be filled by sentences. But one should not assume that this is coherent, or that it can be done without explicit use of the notion of truth. For example, one natural way to specify how a quantifier along these lines works in a sentence of the form "for all s, … s … " is to say that it claims that whatever sentence you put in the position marked by s, the result is a truth.[10]

Paul Horwich (1990) has used T-equivalences (or more exactly ones of the form "the proposition that p is true iff p") to suggest that this is all one needs to say about truth, which is therefore a "non-substantive" property. Such minimalist or deflationary views dominated discussions of truth in the last decade of the last century. (The debate is well represented in Blackburn and Simmons 2000; see also Wright 1992.) Supposedly, all the things we think are important about truth, its relation to objectivity, its value, the fact that we should aim for it, can be derived from reflection on the equivalences.

One well-known reason for thinking that this needs some qualification is that (T) as it stands leads to paradox: far from being trivial, T-equivalence threatens to generate contradictions (see §3.2). Threat of paradox was at the heart of Tarski's work, which constituted the inauguration of formal semantic theorizing, and has cast its shadow over all subsequent serious discussion of truth.

---

[10] Drawing on work by Prior, some recent discussions suggest alternative approaches to sentence quantifiers: see Williamson 1999, Künne 2003.

These truth-related paradoxes have a long history. They were known in antiquity, were studied in the Middle Ages by philosophers like Buridan, and have been the subject of intensive research within philosophical logic for the last hundred years. An early development of this recent phase dates back to Russell, who linked the truth-related paradoxes closely to the class paradox that he discovered in the very early years of the twentieth century and which bears his name (see §3.5). [11] He reported what we now call Russell's paradox in a letter to Frege, who replied in a letter dated 22 June 1902: "Your discovery of the contradiction has surprised me beyond words and, I should almost like to say, has left me thunderstruck, because it has rocked the ground on which I meant to build arithmetic" (Frege 1902: 132). Both Frege and Russell espoused the logicist thesis that mathematical concepts can be defined in purely logical terms, using in particular the theory of classes, and mathematical truths proved from purely logical axioms. This thesis becomes unattractive if the supposedly purely logical theory of classes generates paradoxes, for then one would be reducing apparently unparadoxical mathematics to a paradox-ridden basis.

Russell (1908) provided what he took to be a unified solution to both Russell's paradox and the truth-related ones, in the form of his Ramified Theory of Types, which is designed to make certain kinds of circularity and self-reference impossible, and self-reference and/or circularity were cast as the main villains for most of the first half of the twentieth century. More recently the role of context (see §3.4) has become increasingly significant in this area of philosophy, as it has in others.

The "new logic" was supposed to give thought wings, so it must have been distressing for the founder of philosophical logic to discover that it was a breeding ground for paradox. Russell no doubt found comfort in the view that the Class paradox and the truth-related paradoxes are at bottom the same, for then the new logic did not, after all, breed a new species of monster, but merely allowed existing monsters to don new colours, unless due hierarchical precautions are taken. Whether or not he was right to make that assimilation is considered in §3.5 below.

3.2   The Liar: levels

One version of the Liar is

     L   L is not true.

Here a sentence, called L, is supposed to say of itself that it is not true. Suppose it is true; then it is as it says it is, that is, not true. Suppose that it is not true. This is just what it says it is, and a sentence that correctly says how things are is true, so it is true. To sum up: if it is true it is not true, and if it is not true it is true. This is tantamount to a contradiction.

Using Tarski's schema (T), the contradiction can be derived by using the name "L" in place of "X" and "L is not true" in place of "p" yielding:

     *   L is true iff L is not true.

By our stipulations, "L" meets the condition (imposed by Tarski on the relation between "X" and "p" in instances of (T)) of being a name of "L is not true". So the Liar shows that the apparently platitudinous (T) leads by apparently correct reasoning to the contradictory (*).

Early systematic attempts to deal with such contradictions involved hierarchies of levels. Russell's, as we saw, was a dual-purpose hierarchy which also controlled Russell's paradox. Because of the complexity of Russell's ramified hierarchy, we do better to consider the idea as it was used by Tarski (1933) nearly 30 years later.

---

[11] Class-theoretic paradoxes were known to Cantor as early as 1895 (cf. Kneale and Kneale 1962: 652). Russell's first published account of his paradox is in 1903.

Tarski's response was that a precise account of truth can only be given for formalized languages (ones with a precisely specified grammar, or, to keep closer to Tarski's own formulation, one whose construction ensures that the sense of every expression is uniquely determined by its form), and that Liar-like paradoxes make it mandatory that the metalanguage (the one in which "true" is defined) should be essentially richer than the object language (the one containing the sentences to which the predicate "true" is to be applied).

Suppose some language λ_0 contains a predicate "Tr_1" that applies to all and only the true sentences of λ_0. Suppose also that λ_0 contains a sentence σ that says of itself that it is not Tr_1. Then, granting T, we have a version of the Liar: if Tr_1 applies to σ, then σ says truly that Tr_1 does not apply to it; but if Tr_1 does not apply to it, then, since this is what it says, it is true, and so Tr_1 does apply to it. Tarski took the contradiction to refute the supposition that σ belongs to λ_0. The natural explanation of how this could be is that Tr_1 is not an expression of λ_0. Hence, no sentence belongs to λ_0 if it contains Tr_1. This blocks the paradox in the following sense: the proposed language, since it does not contain a predicate true just of its true sentences, is one in which the paradoxical sentence cannot be formulated. One can write down the words, but they will be devoid of significance: they are semantically wholly defective.

We can enlarge a language by adding new expressions. In particular, we could enlarge λ_0, taken to contain no occurrence of "Tr_1", by adding "Tr_1". We could call the newly formed language λ_1: it contains all the sentences of λ_0, together with all sentences which can be formed from these by using "Tr_1"; so it contains σ. Paradox is still avoided: σ does not belong to λ_0, and since Tr_1 is defined only for λ_0 sentences, there is no question of Tr_1 applying to σ. The expression σ (= "σ is not Tr_1") does not belong to λ_0, and so it is not one of which "Tr_1" can be significantly affirmed or denied.

The construction can continue. There is a predicate, call it "Tr_2", true of just the sentences of λ_1, but it cannot belong to λ_1. In general, a predicate Tr_n cannot belong to a language λ_{n-1} but only to a language of level at least n.

No paradoxical Liar sentence can be formulated in any of the languages in Tarski's hierarchy. How is this supposed to provide a "solution" to the paradox? The paradox arises in our language, so a proper defusing of it must say something about our language, and not merely offer a replacement. In earlier work, Tarski (1933: 153) seems to suggest that the Liar shows that our ordinary language is incoherent. Later, he says that "the semantic conception does conform to a very considerable extent with the commonsense usage" (Tarski 1944: 74). In any case, he is firmly committed to the following: ordinary language is not "semantically closed", where a semantically closed language is one which contains a genuine truth predicate and which can refer to any of its sentences. Yet it seems that our ordinary language is, in this sense, semantically closed. There is no doubt that Tarski's position involves some departure from what we are intuitively inclined to think.

Wherever we encounter paradox, some such departure from intuitive ways of thinking is mandatory. Even so, it is tempting to see how small one can make the departure. One idea in the present connection is that our ordinary language is implicitly a hierarchy of languages, none of which is semantically closed. The plethora of distinct unclosed languages generates an illusion of a single closed one.

A difficulty with this suggestion is that there would appear to be nothing in our usage that reflects the appropriate sensitivity to Tarski-style, fixed-in-advance levels. For example, suppose I say:

> What you said just now is not true.

On the face of it, anyone, including myself, could quite well know what I have said without knowing what you have said. (Imagine a game on the lines of paper, stone, and scissors, in which two players have to make a simultaneous declaration. The task of one is to say whether the other has declared something true. Normally things work well: you declare "Snow is white" and I declare "Not true!" and you win. But what happens when I declare "Not true" and you declare "You win"? My declaration is intelligible in advance of

knowing the content of yours.) On a hierarchical view in which levels are fixed in advance, something in my use of this sentence determines an association between "true" and some level. Presumably the normal (default) level would be 1. If you have said "Snow is white", there is no problem. But suppose you have said "What M. S. will say is true". On the present theory, the intelligibility of my utterance requires my "true" to be on a higher level than yours; but if my utterance can be understood without knowing what you have said, its level of truth must get fixed independently of the content of what you have said. This suggests that we cannot successfully apply this kind of hierarchy response to natural language in this way. (However, compare Burge 1979.)

3.3 *Semantical defects: self-reference, circularity, Revenge*

Responses to the Liar in terms of levels entail that L is semantically defective, that is, does not say anything, true or false, because it does not respect the syntactical constraints which the levels impose. The idea that L is semantically defective can be implemented in other ways. For example, many philosophers in the twentieth century claimed that what is wrong with L is that it refers to itself. This idea, at least in any simple form, is incorrect and inadequate.

It is incorrect because a sentence can refer to itself, as for example this very sentence does, without leading to any kind of semantic defect or paradox. Sentential self-reference cannot be the whole source of Liar paradoxes (cf. Barwise and Etchemendy 1987: 15–6). It is inadequate because one can construct Liar paradoxes without using any sentence which refers to itself. One example of this phenomenon involves Liar cycles like the following.

(A)   (said by α on Monday): Everything β will say on Tuesday is true.

(B)   (said by β on Tuesday): Nothing α said on Monday is true. (Cf. Burge 1978: 90; he attributes the example to Buridan.)

If α and β said nothing other than, respectively, (A) and (B) on, respectively, Monday and Tuesday, we have a paradox of essentially the Liar type. Suppose (B) is true; then (A) is not true, and β will say something not true on Tuesday. Since β only says (B), (B) is not true. So if (B) is true, then it is not true. Suppose (B) is not true; then α said something true on Monday. Since α only said (A), (A) is true, that is, everything β will say on Tuesday is true. This includes (B), so (B) is true. Thus if (B) is not true, it is true.

Neither of the sentences in the story literally refers to itself. Rather, there is a kind of circle, so perhaps we should talk of "circular reference" rather than self-reference. As the circularity doesn't strictly involve reference at all, but rather quantification, it might be safer still just to speak of circularity. Many have seen some kind of circularity at the root of the paradoxes (including Russell); but what kind, exactly?

Russell's idea was that both truth-related and class-related paradoxes arise through infringing what he called the "Vicious Circle Principle" (VCP), one formulation of which is:

VCP   No totality can contain members fully specifiable only in terms of itself.

(Cf. Russell 1908: 75, 63; Russell and Whitehead 1910–13: 37.[12])This is intended as a general metaphysical principle, applicable to classes as well as everything else, and thus also applicable to propositions. The application to classes is straightforward. The Russell class, R, of classes which are not members of themselves, is defined as follows (see also §3.5 below):

for anything y, y ε R iff ¬ (y ε y).

In Russell's eyes, the quantification "for anything, y" introduces a totality, the totality of things which are values of the variable. Suppose this totality contains R. Since the only proper specification of R is the

---

[12] Russell may have borrowed the idea of a vicious circle from Poincaré: cf. Russell and Whitehead 1910: 37n).

displayed sentence just given, this means that the totality would contain a member, R, fully specifiable only in terms of the totality itself, and this is forbidden by VCP. So the totality introduced by "anything" does not contain R, and in that case the instantiation step to the contradictory "R ε R iff ¬ (R ε R)" does not follow.
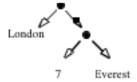
The application to propositions is less straightforward, but is supposed to work along these lines. Consider the totality whose only member is L. How can we fully specify L? Only by identifying it as "L is not true". So this totality can be fully specified only in terms of itself, so there is no such totality. In this argument, we have to think of L not as a sentence, for that clearly exists and is before our eyes, but rather as the proposition the sentence labelled "L" expresses. In effect, VCP is supposed to entail that the sentence in question expresses no proposition.

From the nineteen seventies there was a revival of interest in trying to justify VCP and to show that it rules out all and only the things we want ruled out. One difficulty for it is the apparent consistency of non-foundational set-theory, which allows a class to be a member of itself, in defiance of VCP.[13] Another difficulty is that there appear to be Liar-like paradoxes which do not even involve circularity, let alone self-reference (cf. Yablo 1993, Sorensen 1998).

Another important line of thought about the Liar springs from the idea that truth must depend upon something outside itself: it needs to be "grounded".[14] If "snow is white" is true, this is because snow is white. Our applications of truth and falsehood are intelligible only to the extent that we can find a matter not involving truth (for example, whether or not snow is white) upon which truth can depend. We cannot do this with Liar-paradoxical sentences, and this explains their semantic defectiveness.

Although work inspired by the thought that Liar-sentences are semantically defective has been important philosophically and technically, any view of this kind faces the problem that anything which is semantically defective is not true. On such views, this means that L, being semantically defective, is not true; in short, it means that L is not true. But we have just used a supposedly semantically defective sentence, L itself, to say something we believe to be true. This argument, sometimes called the Liar's Revenge, or Revenge reasoning, needs to be addressed by any adequate treatment of Liar-paradoxes.

---

[13] In non-foundational set-theory, sets are thought of in terms of membership diagrams. The set consisting just of London and the set whose members are the number 7 and Mount Everest (conventionally written: {London, {7, Everest}}) can be represented by the diagram:

Here each blob represents a set, and the branches beneath it represent its members. The following diagram would then represent the set α whose only member is α:

The theory of such diagrams is demonstrably consistent relative to classical set theory. A general account of the evils of circularity, like Russell's VCP, would do well to try to break the link between the diagrams and sets.
Non-foundational set theory owes a great deal to Aczel (1987). A good account, well adapted to present concerns, is by Barwise and Etchemendy (1987), chapter 3.
[14] Classic texts are by Hertzberger (1970) and Kripke (1975).

3.4   *Indexicality and circularity*

We must distinguish between sentences, regarded as things which can be uttered by different people and on different occasions, and the things which people can say or express by using sentences, which I will call statements.[15] One reason to make the distinction is the "indexicality" of language: the fact that the same words may, without ambiguity, be used on different occasions to make different contributions to what is stated, pronouns being a conspicuous example. If you utter the sentence "I am hungry" you make the statement that you are hungry, whereas if I utter the same sentence I make the different statement that I am hungry. The different statements may have different truth values.

Indexicality is at least sometimes at work in Liar-paradoxes. We could expand the story of (A) and (B) by imagining a third utterance:

(C)   (said by γ on Tuesday): Nothing α said on Monday is true.

The fact that β and γ use the very same sentence, yet only one of them is circular in the relevant way, shows that paradox-relevant circularity is not a property of sentences as such. Being meaningful or meaningless is a property of sentences. Since there is nothing paradoxical about (C), there is no reason to say it is other than meaningful, and since (B) is the same sentence, it follows that the property which circularity prevents is not that of being meaningful. We need a more refined notion, one sensitive to the use to which a sentence is put on a specific occasion.

Let us assume that bivalence holds for statements: every statement is either true or false. A sentence can be meaningful, yet on a specific occasion be used in such a way as to fail to make a statement (for example, "That elephant is about to charge", said when no elephant is present). Although sentences can be self-referential, or more generally can have the kind of circularity associated with paradox, it may be that statements cannot. Thus, reverting to the example in the previous section, we might be able to justify the claim that whereas both β and γ use the same sentence, only γ thereby succeeds in making a statement. To suppose that β made a statement would be to suppose that some statements are circular in the exemplified way, and the idea is that no statement can be that.

Once we have the distinction between sentence and statement, we need to recast the Liar paradox. Previously, "L" was supposed to label a sentence, but on the present view, "true" cannot be meaningfully applied to a sentence (as opposed to a statement). This means that there is no intelligible supposition that L is true, and this part of the argument to contradiction would break down. We can revive the paradox in the form:

L*   L* does not express a true statement.

Following the original reasoning, suppose L* does express a true statement. Then things are as it says, so L* does not express a true statement. Suppose it does not express a true statement. Since this is just what it expresses, it does express a true statement. If we say that L* is semantically defective, we immediately generate Revenge reasoning. Since what is semantically defective does not express a statement at all, a fortiori it does not express a true statement; so L* does not express a true statement. We find ourselves trying to use a sentence which, according to our theory, does not express a true statement, in an attempt to express something we take to be true.

By working on the distinction between sentence and statement it may be possible to make sense of this, for a sentence may in one context not be capable of being used to make a statement at all ("that elephant is about to charge", said in the absence of any elephant), even though in another context it can be used to make a true statement. If this applies to the relevant uses of L*, Revenge reasoning will be neutralized.

The general feasibility of such an approach is suggested by cases like the following. Suppose that the following sentence is the only sentence written on the board in room 101:

---

[15] Following Strawson's usage (1950a).

The sentence written on the board in room 101 does not express a true statement.

It appears perfectly consistent for me to write on this page that, because of some semantic defect, the sentence written on the board in room 101 does not express a true statement. I use the words which, as written on the board in room 101, are defective, in circumstances in which there is nothing defective about their use. This suggests that the same words, used to refer to the same thing, and applying the same predicate to it, do not necessarily make the same statement. The sentence written on the board in room 101 makes no statement, in its use in room 101; whereas I use those words, on this page, to make a true statement. I did not have to use the same words. Under suitable circumstances I could simply have said "That sentence does not express a true statement". That there are special circumstances under which I can re-use the same words is an accident of our use of language, and does not affect the truth of the statement I wish to make.

After consideration of the paradox generated by L*, the theorist wishes to say that it is semantically defective and so say that *L* does not express a true statement*. The theorist must hold that his use of the italicized words does express a statement, and a true one, and thus differs from the original use of these words. He could have made his point by using the non-paradoxical words "The sentence displayed above does not express a true statement". It is an accident of our labelling conventions that the very words to be condemned can be used to condemn them. In its defective use, L* calls for the impossible, the existence of a self-referential statement; in its non-defective use it does not.

Could the ingenious opponent not devise Liar sentences for which this response is ruled out? (For example, what should be said about "No use of this very sentence makes a true statement"?) Even if the answer is No, the theorist clearly owes some more details about exactly what features of the shift in context render a sentence which expresses no statement in the first context express a truth in the second.

Tyler Burge (1979) has contributed to this last problem, suggesting that there are different levels of truth, and which level is at issue is fixed not by the meaning of the sentence, but by the statement it is used to make on a given occasion. This approach based on indexical levels avoids many of the difficulties associated with Tarski's hierarchy, which attaches to sentences rather than to statements.

3.5   Russell's paradox: a close cousin?

Are the class-related and truth-related paradoxes totally different, or essentially the same? This is a paradigm of a philosophical logical question. Although Russell, as we have seen, thought of them as variations on a single theme, Frank Ramsey (1925: 183–4) urged that the paradoxes are different in kind, and his view has been dominant, at least until relatively recently. He based the distinction on their different subject matter: the logical paradoxes, under which heading he included Russell's paradox, arise from logical notions, like that of class; the semantic paradoxes, under which he included the Liar, arise from semantic notions, like that of truth. Before discussing the disagreement between Ramsey and Russell, let me briefly reconstruct Russell's paradox.

Because we are inclined to think that every intelligible condition determines a class, we are inclined to think that there is a class, R, whose members are just those classes which are not members of themselves. This is the Russell class, and the paradox consists in the combination of the demonstration that there can be no such class with the intuitions which support principles entailing that there is such a class. The basic intuition is something along these lines:

> CE   For every intelligible condition *F*, there is a class *x*, such that: for anything *y*, $y \in x$ if and
> only if *y* satisfies *F*.

This entails that there is such a class as R if there is the intelligible condition *being a class that is not a member of itself*, for then

> There is a class, R, such that for anything *y*, $y \in R$ iff $\neg (y \in y)$

and by instantiation we get the absurd

RP   There is a class, R, such that R ε R iff ¬ (R ε R).

In addition to Ramsey's basis for claiming that Russell's paradox and the Liar are different in kind (the fact that they turn upon different concepts), there are other differences: there is no analogue in Russell's paradox of Revenge reasoning. There is an immediate problem with the idea that there is no statement expressed by L*, namely that it seems to follow that L* does not express a true statement. No such twist is consequent on the assertion that there is no class R. For the class paradox, there is quite widespread agreement on what we need to say – that there is no class R. What is unclear is how this can be justified. For the Liar, it is not clear even what ought to be said, let alone how to justify it.

So there are certainly some differences between the paradoxes. There are also similarities, and I enumerate five.

(1) The Class paradox resembles a paradox about properties, and the Property paradox in turn resembles the Liar. If we think that every intelligible condition determines a property, then there is a property, P, of being a non-self-applicable property (e.g. the property of being a man is non-self-applicable, since no property is a man). But this property, P, is self-applicable iff it is not self-applicable. One dimension of resemblance is with Russell's paradox, and another is with the contradiction that L truly predicates truth of itself if and only if it does not. Where the Property contradiction uses the notion of *not true of*, embedded in the notion of non-self-applicability, the Liar contradiction uses the notion of *not true*, a property that a sentence or statement may possess.

(2) Both the Class paradox and the Liar seem to involve self-reference, or some similar circularity.

(3) The principles appealed to in the derivation of the two paradoxes (CE: For every intelligible condition F, there is a class x, such that: for any object y, y ε x if and only if y satisfies F; and T: X is true iff p) are structurally similar, and appear to play similarly constitutive roles with respect to the intuitive notions of class and truth.

On the side of derivation, the comparison is that the schema

for any object y, y ε x iff y is F

yields a contradiction when x is replaced by a name, say R, for the Russell class, and F by the condition, expressed in terms of this name, that supposedly defines membership for that class, "¬ … ε R". Similarly, the schema

X is true iff p

yields a contradiction when X is replaced by a name, say L, for the Liar sentence, and p by the condition, expressed in terms of this name, that supposedly defines truth for that sentence, "L is not true".

On the side of roles, the comparison is that just as (CE) appears constitutive of our pretheoretical notion of a class, so (T) appears constitutive of our pretheoretical notion of truth. (CE) determines what it is for a class to exist; (T) determines what it is for a truth condition to exist.

(4) Hierarchies have been used in response to both kinds of paradox, beginning with one of the earliest systematic treatments (Russell 1908). It is natural to suppose that we should think of classes as initially constructed out of non-classes, with each constructional step drawing only upon entities which have already been constructed. Likewise, it is natural to suppose that we should think of statements ascribing truth as initially applied to statements free of the notion of truth, with each constructional step drawing only upon statements which have already been constructed.

(5) Russell's classification of the Class paradox and the Liar as of a common kind is based on the claim that they both alike derive from an infringement of the Vicious Circle Principle.

A formal vindication of Russell's claim that the paradoxes belong to a significant common kind has been provided by Priest (1994). This claim, as I have just expressed it, is consistent with these paradoxes also belonging to significantly different kinds.

The Class and Liar paradoxes, like most things, are similar in some respects, dissimilar in others. Classification matters here because of the constraints it imposes on proper responses to the paradoxes. (Many other paradoxes need to find their place in a classification: see Priest 1994.) If two paradoxes are essentially similar, similar in what really matters, then it is proper to respond in essentially similar ways. For example, if the Class paradox calls for a hierarchy of levels, and the Liar is essentially similar, then it too calls for a hierarchy of levels. If an adequate conception of classes should allow non-well-founded classes, e.g. the class α whose only member is α, and the Liar is essentially similar, then our response to it should allow for analogous circularity. Russell's allegedly common solution to the two paradoxes, his Ramified Theory of Types, has a somewhat specious uniformity, since some of its more complex features are motivated by matters relating to the Liar rather than the Class paradox.

To infer that the paradoxes require a uniform solution, we must show more than that there is a single kind to which they belong. (Unless we have some principled way of restricting kinds, for any two things there is a kind to which both belong.) We need in addition to show that this common kind reveals their common essential nature. I doubt, however, whether this can be done quite independently of views about what response is appropriate to each.

4   The future
Philosophical logic played a central role in analytic philosophy in the twentieth century. Was this a passing phase, or is its influence here to stay? A proper answer involves the distinction, noted towards the end of §2, between philosophical logic understood as the Russell-Quine-Davidson programme, and philosophical logic understood as the philosophical investigation of the central concepts of logic, like truth, validity, and entailment. Investigations of the latter kind have a long history, stretching back to Aristotle and beyond. There is every reason to think that they will continue to play an important part in any future work in philosophy. The Russell-Quine-Davidson programme, by contrast, has had a much shorter history, arising as it did in the twentieth century. I suspect that its fading is already discernible. Philosophers may well resort to logical formalizations in order to clarify claims couched in natural language sentences and arguments. They are increasingly reluctant to affirm that such clarifications reveal something otherwise hidden in the nature of natural language sentences or ordinary thoughts. At the same time, the systematic study of language has developed in ways which make the kind of approach which Davidson championed seem both unnecessary and distorting. The hypothesis that ordinary speakers translate natural language sentences into the formulae of first order logic has low initial plausibility. As Strawson presciently stressed (1950a), there is something about the rigidity of logical formulae which makes them a poor match for so flexible an instrument as natural language. I will illustrate with the example of the anaphoric dependence of pronouns.

Consider the following exchange:
   A: We have a new neighbour.
   B: I know. I saw her moving her stuff in yesterday.
   A: She looks nice.
Given just the resources of classical logic, how should one understand the occurrences of "it"? The only option is to see it as a variable bound by an existential quantifier introduced by "a new neighbour". In classical logic, only closed formulae can be properly evaluated as true or false: a formula with a free variable can be true or false of objects, but not true or false absolutely.[16] Applied to the example, this means that either A's first utterance or B's utterance is not truth evaluable, since it will not be represented as corresponding to a closed formula. Suppose that A's utterance is so represented. Then the "her" in B's

---

[16] There is an overall relativization to a model; it leaves the essence of the point unaffected.

utterance will be represented by an unbound variable, so the utterance is represented as not truth evaluable. Suppose B's utterance is represented as closed. If this is possible at all (and the possibility would involve some changes in the standard definitions of closure), it is by representing A's utterance as open, and so as not truth evaluable: it will have a missing right-hand parenthesis, to enable the scope of its existential quantifier to reach to B's "her". This is a completely implausible upshot. In any case, as A's second utterance shows, the conversation may go on indefinitely, and no one could know when to insert that final right-hand parenthesis. Classical notions of variable binding are quite inadequate to represent this kind of anaphoric dependence in natural language.

This observation was, historically, the point of departure for a new approach to language, Discourse Representation Theory or DRT (cf. Kamp 1981, Kamp and Reyle 1993), which explicitly included the aim of doing more justice than could be done classically to the psychological reality of language processing. It is an example of the kind of approach to language in which the dynamics of discourse are not idealized away, but are treated as central (perhaps "dynamic semantics" might be a suitable phrase for the general nature of the approach, though some also use it in a narrower way); such approaches seem to me likely to make the notion of logical form, as exploited in the Russell-Davidson tradition, obsolete. Whereas in the Russellian tradition truth and reference are essential to an explanation of "meaning", dynamic theories describe meaning in terms of the way it forces or permits changes in the informational states of an interpreter as interpretation proceeds.

The structures whereby DRT represents natural language (these include nested boxes, arrows and first order formulae) are not identical with natural language sentences, and there is nothing in the methodology which requires that all interpretively valuable features can be introspectively identified in a way fitted to systematic theory; there is nothing to preclude "null" or "unsounded" elements" as theoretical posits. So are not DRT structures party to the original programme of logical form, the only difference being that these more sophisticated structures replace first order formulae? Should we not see the same methodology at work, but with a change in what is to count as the "logic" in "logical form? Can it be right to see anything more radical than an improvement of detail?

A difference between Russell's approach and the approach exemplified by DRT and other dynamic semantical theories is this: whereas Russell began with the assumption that the structures of thought are the very structures found in first order logic (or, more exactly, in the higher order logic of *Principia Mathematica*), the new approach makes no assumptions of this kind, and introduces its theoretical structures with no aim other than to provide a revealing description of the data of language use. This transformation of aim, and the consequent weakening of the influence of first order logic as a basis upon which to describe natural language, constitutes a definitive movement away from the programme of logical form as implemented by Russell, Quine and Davidson.

Turning to the other interpretation of "philosophical logic", according to which it is the philosophical study of the concepts important to logic (truth, validity, entailment, consistency, and so): no decline in the importance of this theoretical activity can be envisaged, so long as philosophy itself continues to be practiced. It is in the nature of philosophy to reflect on its aims and tools, attempting to free itself of impossible or worthless goals and of ineffective or distorting tools. The assessment of its arguments as valid or otherwise, its claims as true or otherwise, its propositions as related or not related by entailment: these invite reflection on the nature of the assessment itself. What is it for an argument to be valid, a belief to be true, or for propositions to be related by entailment? So long as philosophy is practiced, in the tradition we inherit from Plato and Aristotle, these questions will not go away.[17]

---

[17] Many thanks to two anonymous referees for this volume, who provided helpful comments.

REFERENCES

Aczel, Peter (1987). *Lectures on Nonwellfounded Sets*. Stanford, CA: CSLI Lecture Notes, No 9.

Austin, J.L. (1961). "Unfair to Facts." In his *Philosophical Papers*, 154-174, Oxford: Clarendon Press.

Barwise, Jon and John Etchemendy (1987). *The Liar: An Essay in Truth and Circularity*. New York and Oxford: Oxford University Press.

Burge, Tyler (1978). "Buridan and epistemic paradox." *Philosophical Studies* 34: 21–35.

Burge, Tyler (1979). "Semantical paradox." *Journal of Philosophy* 76: 169–98. Reprinted in Martin, R.L. (ed.) *Recent Essays on Truth and the Liar Paradox,* Oxford: Oxford University Press, 1984, pp. 83–117.

Cargile, James (1970). "Davidson's notion of logical form." *Inquiry* 13: 129–39.

Chomsky, Noam (1957). *Syntactic Structures*. The Hague: Mouton.

Davidson, Donald (1967). "The logical form of action sentences." In Rescher, Nicholas (ed.) *The Logic of Decision and Action*, Pittsburgh: University of Pittsburgh Press. Reprinted in his *Essays on Actions and Events*, Oxford: Oxford University Press, 1980, pp. 105–22.

Davidson, Donald (1970). "Action and Reaction." *Inquiry* 13: 140–8. Reprinted as "Reply to Cargile" in his *Essays on Actions and Events,* Oxford: Clarendon Press, 1984, pp. 137–148.

Dummett, Michael (1973a). *Frege: Philosophy of Language*. London: Duckworth.

Dummett, Michael (1973b). "The philosophical basis of intuitionistic logic". In his *Truth and Other Enigmas* London: Duckworth, 1978: 215–47.

Evans, Gareth (1982). *The Varieties of Reference*. Oxford: Clarendon Press.

Frege, G. (1902). "Letter to Russell (22 June)", reprinted in *Philosophical and Mathematical Correspondence*, edited by Gottfried Gabriel et al., Oxford: Basil Blackwell, 1980.

Hornstein, Norbert (1995). *Logical Form*. Cambridge, Mass: Blackwell Publishers.

Horwich, Paul (1990). *Truth*. Oxford: Basil Blackwell.

Kamp, H. (1981). "A theory of truth and semantic representation". In J. A. G. Groenendijk, T. M. V. Janssen and M. B. J. Stockhof (ed). *Formal Methods in the Study of Language* Amsterdam, Methamatisch Centrum, 1: 277–322.

Kamp, H. and U. Reyle (1993). *From Discourse to Logic*. Dordrecht, Kluwer.

Kneale, William and Martha Kneale (1962). *The Development of Logic*. Oxford: Clarendon Press.

Kripke, Saul (1975). "Outline of a theory of truth." *Journal of Philosophy* 72, 690-716. Reprinted in Martin, Robert L. *Recent Essays on Truth and the Liar Paradox,* Oxford: Oxford University Press, 1984: 53-81.

Künne, W. (2003). *Conceptions of Truth*. Oxford, Oxford University Press.

Lycan, William G. (1984). *Logical Form in Natural Language*. Cambridge, Mass.: MIT Press.

Martin, R.L. (ed.). (1984). *Recent Essays on Truth and the Liar Paradox*. Oxford: Oxford University Press.

Neale, Stephen (1993). "Logical form and LF." In Otero, C. (ed.) *Noam Chomsky: Critical Assessments*, p. 788–838, London: Routledge.

Priest, Graham (1994). "The structure of the paradoxes of self-reference." *Mind* 103: 25-34.

Quine, W.V.O. (1948). "On what there is." *Review of Metaphysics* . Reprinted in his *From a Logical Point of View,* Cambridge, Mass.: Harvard University Press, 1961: 1–19.

Ramsey, Frank (1925). "The foundations of mathematics." *Proceedings of the London Mathematical Society* 25: 338–84. Reprinted in D. H. Mellor (ed.) *F. P. Ramsey: Philosophical Papers,* Cambridge: Cambridge University Press, 1990, pp. 174–224.

Russell, Bertrand (1903). *The Principles of Mathematics*. Cambridge: Cambridge University Press.

Russell, Bertrand (1905). "On denoting." *Mind* 144: 79-93. Reprinted in Marsh, R.C. (ed.), *Logic and Knowledge,* London: Routledge and Kegan Paul, 1956, pp. 41-56.

Russell, Bertrand (1908). "Mathematical logic as based on the theory of types." *American Journal of Mathematics* 30: 222-62. Reprinted in Marsh, R. C. (ed.) *Logic and Knowledge,* London: George Allen and Unwin, 1956, pp. 59-102.

Russell, Bertrand (1912). *Problems of Philosophy*. Since 1959 published by Oxford Univeristy Press (Oxford and New York).

Russell, Bertrand (1913). *Theory of Knowledge*. First pub. in E. R. Eames and K. Blackwell (eds.) (1984), *The Collected Papers of Bertrand Russell, vol. 7: Theory of Knowledge, The 1913 Manuscript*. London: George Allen and Unwin.

Russell, Bertrand (1914). "Logic as the essence of philosophy." In *Our Knowledge of the External World*,

London: George Allen and Unwin, pp. 42–69.

Russell, Bertrand (1959). *My Philosophical Development*. London: George Allen and Unwin.

Russell, Bertrand, and Whitehead, Alfred North (1910-13). *Principia Mathematica*. Cambridge: Cambridge University Press.

Sainsbury, R.M. (2002). "How can some thing mean something?" in his *Departing From Frege*, London: Routledge.

Soames, S. (1999). *Understanding Truth*. Oxford, Oxford University Press.

Sommers, Fred (1982). *The Logic of Natural Language*. Oxford: Clarendon Press.

Strawson, Peter (1950a). "On Referring." *Mind* 59: 269-86. Reprinted in his *Logico-Linguistic Papers,* London: Methuen, 1971, pp. 1-27.

Strawson, Peter (1950b). "Truth." *Supplementary Proceedings of the Aristotelian Society* 24: 129–56. Reprinted in Pitcher, George (ed.) *Truth,* Englewood Cliffs, NJ: Prentice Hall, 1964: 32–53.

Tarski, Alfred (1933). "The concept of truth in formalized languages." Reprinted in his *Logic, Semantics, Metamathematics,* Oxford and New York: 1956: 152-278, Clarendon Press.

Williamson, T. (1999). "Truthmakers and the converse Barcan formula." *Dialectica* 53: 253-270.

Wittgenstein, Ludwig (1921). *Tractatus Logico-Philosophicus*. London: Routledge and Kegan Paul.

Wright, Crispin (1992). *Truth and Objectivity*. Cambridge, Mass: Harvard University Press.

A Bibliographical Guide to the Subject

A good starting point for work in philosophical logic is Bertrand Russell's "On denoting" (*Mind* 14: 79-93, reprinted in Marsh, R.C. (ed.), *Logic and Knowledge,* London: Routledge and Kegan Paul, 1956, pp. 41-56), or alternatively his later and more accessible expression of essentially the same view in chapter 16 of his *Introduction to Mathematical Philosophy* (1919, London: George Allen and Unwin). This is a paradigm of a contribution to philosophical logic, and even though it does not make explicit mention of the idea of logical form, the theory can be rightly thought of as a proposal about the logical form of judgements expressible with definite descriptions. If one's main interest were in the theory of descriptions, one would consult Stephen Neale's excellent *Descriptions* (Cambridge, Mass.: MIT Press, 1990) and the collection edited by Gary Ostertag (*Definite Descriptions: A Reader*. Cambridge, Mass: MIT Press, 1998). If one's main interest were in pursuing the idea of philosophical logic and logical form, one might move forwards in time to Russell's explicit introduction of the term "philosophical logic" (in "Logic as the essence of philosophy", in *Our Knowledge of the External World*, London: George Allen and Unwin, 1914: 42–69); and also, in order to improve the historical perspective, backwards to his *Principles of Mathematics* (Cambridge: Cambridge University Press, 1903) which, though designed to establish the logicist thesis that mathematics is part of logic, contains some detailed and often agonized discussions of such philosophical logical topics as proper names and denoting expressions. It also contains the first published explicit formulation of Russell's paradox.

Russell's conception of logical form almost certainly influenced Rudolph Carnap, who certainly influenced Quine and, thereby, Davidson. In, for example, Carnap's *The Logical Syntax of Language* (London: Routledge and Kegan Paul, 1937) we find a highly Russellian vision of the dissolution of philosophy into either empirical science or logic: "we shall here maintain that all these remaining philosophical questions [all those which remain once all empirical questions have been assigned to the appropriate special science] are logical questions. … once philosophy is purified of all unscientific elements, only the logic of science remains. … *the logic of science takes the place of the inextricable tangle of problems which is known as philosophy*" (p. 279). Carnap's "logic of science" is quite close to Russell's "philosophical logic".

For many years Russell's conception of logical form, and his programme in philosophical logic, escaped explicit challenge (though there was plenty of philosophical activity, even within the Anglo-Saxon tradition, that did not conform to it). A serious attack on many Russellian preconceptions is found in Peter Strawson's "On referring" (*Mind* 59: 269-86 (1950), reprinted in his *Logico-Linguistic Papers,* London: Methuen, 1971: 1-27). The main target is Russell's theory of descriptions, but in the course of the attack the very idea of straightjacketing ordinary language into "logical forms" is challenged, along with the thought that one can do justice to the relevant facts of meaning without taking context of use into account.

By the late nineteen sixties, the methodology of logical form was receiving explicit and not always complimentary attention both from philosophers and from linguists. Donald Davidson's papers "The logical form of action sentences" (in Rescher, Nicholas (ed.) *The Logic of Decision and Action*, Pittsburgh: University of Pittsburgh Press, 1967, reprinted in his *Essays on Actions and Events*, Oxford: Oxford University Press, 1980, pp. 105–2) and "Action and Reaction" (*Inquiry* 13: 140–8, 1970, reprinted as "Reply to Cargile" in his *Essays on Actions and Events,* Oxford: Clarendon Press, 1984, pp. 137–148) are essential reading, as is also the article by James Cargile which prompted the second one ("Davidson's notion of logical form." *Inquiry* 13: 129–39, 1970). Davidsonian approaches have been developed by others, for example William Lycan (*Logical Form in Natural Language* Cambridge, Mass.: MIT Press, 1984), which includes a useful summary sketch of relevant historical developments at pp. 4–9. The doctrine that logical form should reveal all the logically significant features of a sentence has been attacked by John Etchemendy ("The doctrine of logic as form", *Linguistics and Philosophy* 6: 319–334, 1983), who states this aspect of the logical form programme as follows: "sentences with different logical properties [must] be assigned different logical structures" (p. 319). In the mainstream logical form tradition, set by Russell, Quine and Davidson, it was argued, or more often taken for granted, that logical forms would be, or be represented by, formulae of familiar languages (first order languages or their extensions). There is no reason in principle why one should not combine the quest for formal rigor and the supposition that logical forms and surface forms may differ with a very different conception of the underlying logic. This

possibility is impressively fleshed out by Fred Sommers in his *The Logic of Natural Language* (Oxford: Clarendon Press, 1982).

Linguists introduced the notion of logical form, or the LF level of representation, with rather different motivations, and it is a controversial question to what extent linguists' LF coincides with philosophers' logical form. To see the linguists' notion in use, a classic text is R. May's *Logical Form: its Structure and Derivation* (Cambridge, Mass.: MIT Press, 1985), and a more recent one is Norbert Hornstein's *Logical Form* (Cambridge, Mass: Blackwell Publishers, 1995). Some linguists, like Greg Carlson in "Logical form: types of evidence" (*Linguistics and Philosophy* 6: 295–317, 1983) are disposed to start with some skepticism about the necessity for LF in a complete account of language. For an assessment of the relation between logical form and LF see Stephen Neale ("Logical form and LF", in C. Otero (ed.) *Noam Chomsky: Critical Assessments*, London: Routledge, 1993: 788–838), who concludes that logical form in the philosophers' sense coincides with the linguists' LF, the upshot being that "Chomsky's work on the syntax of natural language has application well beyond theoretical linguistics" (827). On the other hand, Daniel Sperber and Deidre Wilson (*Relevance: Communication and Cognition*, Oxford: Blackwell Publishers, 1995) use the notion of logical form for the semantic aspects of an utterance which are recovered in utterance interpretation by an automatic process of decoding. They claim that the logical form may not have complete truth conditions and typically underdetermines the proposition expressed. This usage is in clear conflict with most philosophical conceptions of logical form, which apply to sentences rather than to utterances, and are assumed unambiguously to express a complete truth-condition.

The use of the notion of logical form to derive ontological conclusions faces the difficulty that the original sentence and its logical form are usually supposed to be equivalent in a way that should ensure that they have the same ontology (that is, that the same things need to exist for the one to be true as for the other). If "Shem kicked Shaun" and "There is a kick which Shem gave Shaun" are equivalent, should we say that since the former apparently does not require the existence of kicks, nor does the latter? Or should we say, rather, as Davidson urges, that since the latter does require the existence of kicks, so does the former? W. V. Quine tries to resolve the problem by denying that the equivalence is tight enough to preserve ontology, saying that it is enough that the preferred sentence, the paraphrase, be usable for more or less the same purposes as the one with the dubious ontology (see his *Word and Object*, New York: Technology Press of MIT and John Wiley and Sons Inc., 1960, p. 214). This has not convinced all commentators: see, for example, William Alston ("Ontological commitment", *Philosophical Studies* 98–17, 1958), and for a clear statement of the problem and an unusual solution, Stephen Yablo ("A paradox of existence" in Anthony Everett and Thomas Hofweber (eds.) *Empty Names, Fiction and the Puzzles of Non-Existence*, Stanford, Calif.: CSLI Publications, p. 275–312, 2000, esp. Appendix A).

Possible starting points for work on truth include the following: Susan Haack's *The Philosophy of Logics* (Cambridge: Cambridge University Press, 1978), a book which also provides a good treatment of a number of other themes in philosophical logic; John Mackie's *Truth, Probability and Paradox* (Oxford: Oxford University Press, 1973); Paul Horwich's *Truth* (Oxford: Basil Blackwell, 1990); Crispin Wright's *Truth and Objectivity* (Cambridge, Mass.: Harvard University Press, 1992); the collection of essays edited by Simon Blackburn and Keith Simmons (*Truth* Oxford: Oxford University Press, 1999), which reprints some of the most important papers on this topic from the last hundred years; Scott Soames's *Understanding Truth* (Oxford University Press, 1999), Wolfgang Künne's *Conceptions of Truth* (Oxford University Press, 2003), and Aladdin M. Yaqūb's *The Liar Speaks the Truth: A Defense of the Revision Theory of Truth* (New York: Oxford University Press, 1993), which stresses the significance of the Tarski-equivalences. Horwich offers a deflationist account of truth, according to which it does not amount to a "substantive" property: all we need to know about it is given by equivalences similar to Tarski-equivalences; he also gives useful summary statements and criticisms of other positions. Wright pursues a modified anti-realist line, in which truth is many areas is applicable in a way which fails to guarantee objectivity of the subject matter. Blackburn and Simmons give a useful overview of deflationary or minimalist positions in their editorial introduction, and many of the essays they collect pursue that theme. David Wiggins provides a valuable sketch of connections between truth, meaning, assertion and convergence in his "What would be a substantial theory of truth?" (in Z. van Straaten, (ed.) *Philosophical Subjects: Essays Presented to P. F. Strawson*, Oxford: Oxford University Press, 1980: 187-221). A different approach is taken by Julian Dodd (*An Identity Theory of Truth*, Basingstoke: Macmillan, 2000), who argues that a correspondence theory

cannot give an appropriate account of facts which, as Strawson earlier suggested, are nothing but true thoughts, and this leads Dodd to his ingeniously defended version of an identity theory: to be true is to be identical with a fact.

It is often said that St Paul provides a version of the Liar (epistle to Titus 1, xii–xiii):
>12. One of themselves, even a prophet of their own, said, The Cretans are always liars, evil beasts, slow bellies.
>13. This witness is true. Wherefore rebuke them sharply, that they may be sound in the faith.

It is not clear that the saint sees any logical as opposed to moral problem. For a thirteenth century discussion of many Liar-like paradoxes see John Buridan's *Sophismata* (translated and edited by G. E. Hughes under the title *John Buridan on Self-Reference*, Cambridge: Cambridge University Press, 1982). Russell's account of truth-related paradoxes and their relation to class paradoxes can be found in his "Mathematical logic as based on the theory of types" (reprinted in R.L. Marsh (ed.) *Logic and Knowledge*, London: Allen and Unwin, 1956: 59-102) and in the Introduction to his and Whitehead's *Principia Mathematica* (Cambridge: Cambridge University Press, 1910–13).

A starting point for contemporary discussions of Liar-like paradoxes is Mackie's *Truth, Probability and Paradox* (op. cit.) and the essays in R. M. Martin's collection (*Recent Essays on Truth and the Liar Paradox*. Oxford: Oxford University Press, 1984). For an intriguing monograph, which begins with a very clear overview of the problem, see Jon Barwise and John Etchemendy's *The Liar: An Essay in Truth and Circularity* (New York and Oxford: Oxford University Press, 1987). (This work also contains an accessible introduction to Aczels' theory of non-wellfounded sets.) Alfred Tarski's "The concept of truth in formalized languages" (first published in Polish in 1933 and reprinted in his *Logic, Semantics, Metamathematics,* Oxford and New York: Clarendon Press, 1956: 152-278) fairly quickly becomes dauntingly technical, but the early pages set out clearly the main philosophical motivations. In this paper he asserts: "In [colloquial] language it seems to be impossible to define the notion of truth or even to use this notion in a consistent manner and in agreement with the laws of logic" (p. 153). In the more popular version of some of his ideas in "The semantic conception of truth and the foundations of semantics" (reprinted in A.P. Martinich (ed.) *The Philosophy of Language*, 3rd ed, Oxford: Oxford University Press, 1996: 61–84), he seems to take a more gentle line with ordinary language: "I happen to believe that the semantic conception does conform to a very considerable extent with the commonsense usage" (p. 74). He offers a later popular account in the *Scientific American* article "Truth and proof" (194: 63-77 (1969)). Anil Gupta and Nuel Belnap provide (Section II) a careful statement of Tarski's precise premises, together with a challenge to the full generality of the conclusion Tarski drew, in their *The Revision Theory of Truth* (Cambridge, MA: MIT Press, 1993). For an excellent and accurate account of Tarski's work, stressing in particular the route Tarski saw from the semantic conception of truth to the Gödelian incompleteness of proofs in languages with reasonable expressive resources, see Jeff Ketland's "Deflationism and Tarski's paradise" *Mind* 108: 69–94, 1999. Ketland suggests that the fact that Tarski's theory is in a certain sense non-conservative poses a difficulty for those who see in Tarski's work support for a deflationary or minimalist view of truth. This conclusion needs to be set alongside Yaqūb's claim that a deflationary *conception* of truth suggests that a proper *theory* of truth will be non-deflationary (op. cit.: 42).

For a post-Tarskian attempt to work out a suitable notion of circularity, see Charles Chihara's *Ontology and the Vicious Circle Principle* (Ithaca, NY: Cornell University Press, 1973). Jørgen Jørgensen ("Some reflections on reflexivity", *Mind* 62: 289–300, 1953, reprinted in Bartlett, Steven (ed.), *Reflexivity: A Source-Book in Self-Reference,* Amsterdam and London: North-Holland, 1992) gives an account of the supposed horrors of self-reference. An early source of the notion of grounding is given by Hans Herzberger ("Paradoxes of grounding in semantics", *Journal of Philosophy* 67, 145-67, 1970) and this idea also features in Saul Kripke's "Outline of a theory of truth" (*Journal of Philosophy* 72, 690-716, 1975, reprinted in Martin, Robert L. (ed.) *Recent Essays on Truth and the Liar Paradox,* Oxford: Oxford University Press, 1984: 53-81). Kripke's paper presents a precise theory of truth for a language which both allows circular reference and contains its own truth predicate, a combination which Tarski had supposedly demonstrated was impossible. One alleged difficulty with Kripke's approach is that it is vulnerable to Revenge reasoning. This can be brought out either by considering a Liar-like sentence which purports to say of itself that it is either false or ungrounded, or by the reflection that what is not grounded is, presumably, not true.

Targeting self-reference or circularity as the villains of Liar paradoxes has become less popular in recent years. Among the causes are the appreciation that something like self-reference is used in the construction of various metamathematical proofs, for example Gödel's proof of the incompleteness of arithmetic; a growing appreciation of the need to do justice to our actual language, in which self-reference seems a clear possibility; and the emergence of paradoxes, like those invented by Stephen Yablo ("Paradox without self-reference" *Analysis* 53: 251–2, 1993), which support his conclusion that "self-reference is neither necessary nor sufficient for Liar-like paradox" (p. 252).

The use of indexicality in dealing with Liar paradoxes goes back at least to Haim Gaifman's "Paradoxes of infinity and self-application, I" (*Erkenntnis* 20: 131–55, 1983). A more recent contribution on these lines is offered by Laurence Goldstein in his paper "'This statement is not true' is not true" (*Analysis* 52.1: 1-5, 1992). Tyler Burge suggests that what indexicality should do is generate context-sensitive levels: see his "Semantical paradox" (*Journal of Philosophy* 76: 169–98, 1979, reprinted in Martin, R.L. (ed.) *Recent Essays on Truth and the Liar Paradox,* Oxford: Oxford University Press, 1984: 83–117).

Graham Priest's *In Contradiction* (Dordrecht: Nijhof, 1987) has a clear introduction to both semantic paradoxes (chapter one) and paradoxes involving classes (chapter two). It also offers an unusual solution: some contradictions are true, and examples may include the Liar and the claim that Russell's class exists. For an introductory text on type theory, see Irving Copi's *The Theory of Logical Types* (London: Routledge and Kegan Paul, 1981).

Comparing the Liar and Russell's paradoxes requires noting Frank Ramsey's opinion that "It is not sufficiently remarked, and the fact is entirely neglected in *Principia Mathematica*, that these contradictions fall into two fundamentally distinct groups …"("The foundations of mathematics" *Proceedings of the London Mathematical Society* 25: 338–84, 1925, reprinted in D. H. Mellor (ed.) F. P. Ramsey: Philosophical Papers, Cambridge: Cambridge University Press, 1990, pp. 174–224, at p. 183). Russell's alleged neglect is manifest in the Introduction to his and Alfred Whitehead's *Principia Mathematica* (Cambridge: Cambridge University Press, 1910; the early parts are now reprinted in paperback). The most significant recent contribution is Priest's "The structure of the paradoxes of self-reference" (*Mind* 103: 25-34, 1994).

Discourse Representation Theory dates to Hans Kamp's "A theory of truth and semantic representation", in J. A. G. Groenendijk, T. M. V. Janssen and M. B. J. Stockhof (ed). *Formal Methods in the Study of Language* Amsterdam, Methamatisch Centrum, 1, 1981: 277–322. A more detailed source (the "blue bible") is Kamp and Reyle's *From Discourse to Logic* (Dordrecht, Kluwer, 1993). An approach with some similarities to Kamp's was developed independently by Irene Heim (1982/88): *The Semantics of Definite and Indefinite Noun Phrases*. Ann Arbor, University Microfilms International.

For an impression of the richness of the topics that can be properly counted as belonging to philosophical logic, see a collection by David Lewis, one of the outstanding philosophers of the twentieth century. His *Papers in Philosophical Logic* (Cambridge: Cambridge University Press, 1998) discuss, among other things, adverbs, context, conditionals and probabilities, intensional logic, verificationism, set theory, and mereology.

Biographical sketches:

Davidson, Donald (1917–2003): student of Quine's at Harvard in the 1940s, Davidson was led into questions of philosophical logic by his early reflections on the nature of decision and action. The connection is marked by one of his most important papers on the program of philosophical logic "The logical form of action sentences". Among his main theories are his approach to meaning through the notion of radical interpretation, and his claim that mental and physical properties are "anomalous", that is, not connected in law-like ways. He died unexpectedly when still very active, with a heavy schedule of lectures and conferences.

Dummett, Michael (1925–): Oxford philosopher, best known for his anti-realism and for his studies of Frege. Knighted in 1999 for his services to philosophy. Also important for his work against racism and his studies of tarot cards and other games.

Frege, Gottlob (1848–1925): German mathematician, logician and philosopher, often regarded as the father of analytic philosophy. Best known for his innovations in logic (especially his invention of the modern system of quantification) and his attempt to establish logicism (the view that mathematics reduces to logic), an attempt which he regarded as having failed when Russell pointed out that contradictions could be proved in his logical system.

Quine, Willard van Orman (1908–2000): American philosopher and logician, a dominant figure in the sixties and seventies. Best known for his attacks on logical positivism and the analytic/synthetic distinction, and for moulding an account of a sparse ontology which satisfied his love of "desert landscapes".

Russell, Bertrand (1872–1970): British philosopher, logician and mathematician, best known for his *Principia Mathematica* (with A. N. Whitehead), a milestone in the logicist program and a work with wide-ranging philosophical implications; and for developing a metaphysical system he called "logical atomism". In addition to his strictly philosophical activities, he was a popular writer (treating topics such as marriage, sex and happiness), was awarded the Nobel prize for literature in 1950, and was imprisoned for a week on account of his anti-nuclear protests in 1961.

Strawson, Peter (1919–): Oxford philosopher, who made his name with the article "On referring" (1950) which attacked Russell's theory of descriptions. Associated with "ordinary language philosophy", and known for subtle and elegant positions on a wide variety of topics. Knighted in 1977.

Tarski, Alfred (1902–1983): Polish born logician, emigrated to America in 1939. Laid the foundation for systematic semantic studies, and, along with Russell and Gödel, is arguably one of the three greatest logicians of the twentieth century.

Wittgenstein, Ludwig (1889–1951): Austrian born philosopher, worked with Russell and Moore in Cambridge in the decade before the First World War, wrote the famous *Tractatus Logico-Philosophicus* while a prisoner of war, was a schoolteacher for a decade after the war, and then returned to Cambridge. Only the *Tractatus* was published in his lifetime. The probably more important *Philosophical Investigations*, along with various other works, were published posthumously.

Biographical note:

R. M. Sainsbury was born in London in 1943. He is Professor of Philosophy at the University of Texas at Austin and Susan Stebbing Professor of Philosophy at King's College London. He is the author of *Russell* (London: Routledge, 1979, 1985) *Paradoxes* (Cambridge: Cambridge University Press, 1988, second edition 1995) and *Logical Forms* (Oxford: Blackwell, 1991, second edition 2000), *Departing From Frege* (London: Routledge, 2002) and *Reference Without Referents* (Oxford: Oxford University Press, 2005).