6

Truth

Ian Rumfitt

Some people may doubt if there is anything left for a philosopher to say about truth. We have a practical grasp of the difference between 'true' and 'false' from childhood, and Aristotle gave a nice theoretical account of the notions nearly two-and-a-half thousand years ago. 'To say of what is that it is not', he wrote, 'or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true' (*Metaphysics* Γ 1011b25). There are, though, different ways of developing Aristotle's idea. In this paper, I want to elaborate a natural but curiously neglected version of it and show how the result illuminates both the Paradox of the Liar and the 'axiomatic' approaches to truth that many logicians now favour. My preferred elaboration of Aristotle is certainly not novel, but I have not seen it developed in quite the way that I shall recommend. As usual in philosophical logic, the precise articulation matters.

6.1 Ramsey's account of truth

'To say of what is, that it is, is true'—or, as one might more naturally put it in English, to say so much is to speak truly. In speaking truly, or falsely, the speaker will have produced a declarative utterance (which I understand to include written or printed inscriptions, as well as instances of speech). Some people deem it infelicitous to predicate truth or falsehood of such utterances, as opposed to the propositions or thoughts that they express. One may, though, say that, in producing a declarative utterance, its speaker spoke truly (or spoke falsely), and precisely this is what I shall mean by calling a declarative utterance true (or false).

Ian Rumfitt, *Truth*. In: *Oxford Studies in Philosophy of Language Volume 1*. Edited by Ernie Lepore and David Sosa, Oxford University Press (2019). DOI: 10.1093/oso/9780198836568.003.0006

Is there a way of parlaying Aristotle's *dictum* into definitions of truth and falsity for declarative utterances—or at least, into necessary and sufficient conditions for such an utterance to be true, or false? The *dictum* certainly yields a gloss on the notions of speaking truly and speaking falsely, when we are given what the speaker says. As many writers have remarked, though, it does not provide a general elucidation of truth, even as the notion applies to declarative utterances, for we often apply truth to utterances when what they say is not given. What we would like to find are statements of the form 'Utterance *u* is true if and only if . . . *u* . . . ' and '*u* is false if and only if . . . *u* . . . ' which apply generally, not merely when we are given the content of *u*. Some may be tempted to say: '*u* is true if and only if *u* expresses a true proposition' but that proposal just pushes the problem back. For we now have to say what it is for a proposition to be true—and also what propositions are.

I shall not canvass the diverse proposals that philosophers have made in their quest for general elucidations of truth and falsity of the kind sought. Rather, I want to focus on the suggestion which seems to be the most promising. In *On Truth*, a book he drafted in 1927–9 but which only appeared some sixty years after his death (Ramsey 1991), Frank Ramsey took the primary bearers of truth and falsity to be *beliefs* or *judgements*—terms he used synonymously to apply to particular mental states which have both propositional content and 'some degree of the affirmative character' that is present in judging that such-and-such is the case but absent from merely wondering whether it is the case (Ramsey 1991: 8). He then proposed a definition of truth as it applies to beliefs in this sense:

> Any belief whatever we may symbolize as a belief that *p*, where '*p*' is a variable sentence just as '*A*' and '*B*' are variable words or phrases (or terms as they are called in logic). We can then say that a belief is true if it is a belief that *p*, and *p*. In Mr Russell's symbolism
>
> *B* is true :=: ($\exists p$). *B* is a belief that *p* & *p*. Df
>
> (Ramsey 1991: 9, incorporating 15, n. 7).

Similarly, and in an updated symbolism, *B* is false if and only if ($\exists P$)(*B* is a belief that $P \wedge \neg P$) (Ramsey 1991: 11). Parallel accounts may be given for other sorts of truth bearer. I have been calling a declarative utterance true when, in producing it, its speaker speaks truly. In the same spirit, let us deem such an utterance to *say that P* when, in producing it, its speaker thereby expresses the thought that *P*. (Thus an utterance which says that *P* need not have the force of an assertion.) Given this notion of

saying, we may adapt Ramsey's account to provide elucidations of truth and falsity for individual utterances in the form sought:

(T_R) u is true if and only if $\exists P(u$ says that $P \wedge P)$

(F_R) u is false if and only if $\exists P(u$ says that $P \wedge \neg P)$.

Versions of this account (which is different from the better known 'redundancy theory' that Ramsey had propounded earlier[1]) remain popular. Timothy Williamson uses 'True (s, c)' to mean 'the [type] sentence s is true as uttered (with its present meaning) in the context c' and 'Say (s, c, P)' to mean 'In c, s expresses the proposition that P' (Williamson 1998: 7, 12). He then proposes 'explicit definitions of truth and falsity' that conform to Ramsey's template:

(T_W) True $(s, c) \equiv \exists P(\text{Say } (s, c, P) \wedge P)$

(F_W) False $(s, c) \equiv \exists P(\text{Say } (s, c, P) \wedge \neg P)$ (ibid. 14).

We may think of (T_W) and (F_W) as extending the scope of the Ramseyan accounts of truth and falsity to encompass all the possible utterances that could be made by uttering various type sentences in various contexts.

Both Ramsey and Williamson describe the equivalences they propose as *definitions* of truth and falsity. It is not altogether clear what they mean by this, and it is doubtful if any of these equivalences could serve as a definition in the strong sense of enabling someone who had not even an implicit grasp of the term being defined to understand it. For the definitions take as understood the notion of an utterance's saying that P, i.e. expressing the thought that P, and it is plausible to hold that part of understanding what it is for an utterance to say that P is knowing that it is true if and only if P. If that is right, then the Ramseyan equivalences do not define truth and falsity in more basic terms. At best, they contribute to the definition of a package deal of interrelated concepts, somewhat as Newton's Laws of Motion contribute to the definition of force and inertial mass. Even if we disclaim all reductive or definitive ambitions, though, we can still ask if Ramsey's equivalences correctly characterize the relationship between the notions of truth, falsity, and content, and it is this question which will be my focus here.

[1] See Rumfitt 2011 for an analysis of the differences.

In addressing it, one urgent task is to explicate the existential quantifier '$\exists P$' that appears in all four equivalences. How is this to be understood?

It is certainly not to be read as an objectual quantifier. Objects, in the logical sense, are things that might be signified by singular terms, but the variables bound by '$\exists P$' replace complete sentences, not terms. Neither, though, is '$\exists P$' to be read as a substitutional quantifier. The substitution class associated with a substitutional quantifier 'ΣP' can certainly include complete sentences, but '$\Sigma P(u$ says that $P \wedge P)$' means 'There is a sentence S in the substitution class which yields a truth when it is substituted for "P" in "u says that $P \wedge P$"'. While this meaning is clear, 'yields a truth' presupposes an antecedent understanding of what it is for an utterance or inscription to be true. Accordingly, '$\Sigma P(u$ says that $P \wedge P)$' can provide no elucidation of 'u is true'.

It was one of Arthur Prior's insights that these alternatives do not exhaust the available readings of '$\exists P$'. Natural languages, he remarked, contain a gamut of non-nominal quantifiers: in English we have among others 'however', 'somehow', 'wherever', and 'somewhere' (Prior 1971: 37). 'No grammarian', Prior comments,

> would count 'somehow' as anything but an adverb, functioning in 'I hurt him somehow' exactly as the phrase 'by treading on his toe' does in 'I hurt him by treading on his toe'. (ibid.)

As he further observed, non-nominal quantifiers like these help us to interpret the formal devices '$\exists P$' and '$\forall P$'. '$\exists P$' may be read as 'things are somehow', '$\forall P$' as 'however things may be', and 'P' as 'things are thus'. So '$\exists P(u$ says that $P \wedge P)$' may be understood as saying: 'u says that things are somehow, and they are thus'; this is an initially plausible elucidation of 'u is true'. Similarly, '$\exists P(u$ says that $P \wedge \neg P)$' says 'u says that things are somehow, and they are not thus'; this captures 'u is false'. Prior's gloss on '$\exists P$', then, provides a nice English rendering of Ramsey's account of truth and falsity. It also allays any suspicion that might arise from the same bound variable's appearing in an intensional context (following 'u says that') and in an extensional context (e.g. after the conjunction sign). A way things might be may be *both* the way someone says things are on particular occasion *and* a way things actually are.[2]

[2] On this interpretation, there is a mismatch between the substituents of the variable 'P' and the entities over which '$\forall P$' and '$\exists P$' are understood to range. Occurrences of 'P' will be

When so interpreted, sentential quantifiers also make possible an attractive way of stating logical laws. Under the influence of Quine, many philosophers imagine that such laws are inherently schematic; thus, to accept the Law of Excluded Middle is to accept every instance of the schema ⌜Either *A* or not *A*⌝. In this formulation, the logical words 'either . . . or' and 'not' are mentioned, not used, and while there will be corresponding schemata in languages other than English, the fact remains that on this approach logical laws are tied to their linguistic expression. Sentential quantifiers, under Prior's interpretation, help formalize an alternative conception. We may formulate the Law of Excluded Middle as 'However things may be conceived to be, either they are thus or they are not thus', and put this into formal dress as '$\forall P(P \vee \neg P)$'. The variable '*P*' now ranges, not only over the ways things might (metaphysically) be, but over all the ways things could be conceived to be. (Different logical schools will differ over which ways fall within this range.) In these formulations, the connectives are used, not mentioned, which has advantages when it comes to the modal status of logical laws. It is a wholly contingent matter that the English word 'either' (or the formal symbol '$\vee$') signifies disjunction, so it is a merely contingent fact (if it is a fact at all) that every instance of the schema ⌜Either *A* or not *A*⌝ is true. If, however, Excluded Middle is a logical law, then it will be true to say 'It is logically necessary that however things may be conceived to be, either they are thus or they are not thus', i.e. '$\Box\forall P(P \vee \neg P)$'.

replaced by complete well-formed formulae, whereas ways things might be are signified by adverbial clauses. This mismatch reflects profound grammatical differences between English and familiar formal languages and I do not think it prevents Prior's glosses from capturing the intended senses of formulae containing sentential quantifiers. If we wanted to, we could take a leaf out of Frege's first book, the *Begriffsschrift*, and regard e.g. '$\exists P\Phi(P)$' as an abbreviation of '$\exists P$—$\Phi(P)$', with the horizontal sign '—' understood as 'Things are', and '$\Phi(P)$' construed as an adverbial clause. Indeed, we could apply this way of reading formulae across the board, so that the specific content of every well-formed formula is localized in an adverbial clause which always combines with 'Things are' to yield a complete sentence. (Compare the way Frege localized the specific content of each well-formed formula of his *Begriffsschrift* in a singular term which always combines with the predicate 'is a fact' to yield a complete sentence; see Frege 1879: §3.) Such an interpretation would result in no substantial deviation from the customary use of the formalized language. (I am here indebted to discussion with Christopher Peacocke.)

6.2 Strawson's account of truth

Prior's glosses do not fully explain '$\exists P$' and '$\forall P$'. They give a rough indication of the intended sense of these quantifiers but their logical properties are not thereby determined. There is, though, enough on the table to introduce a rival account of truth and falsity. This rival belongs to the same family as Ramsey's. It, too, characterizes truth and falsity in terms which take as understood the notion of an utterance's saying that *P*, or a belief's being a belief that *P*. It also uses sentential quantifiers.

The *locus classicus* of the account I have in mind is Peter Strawson's Inaugural Lecture at Oxford, 'Meaning and Truth' (Strawson 1970). There he proposed, as 'something uncontroversial and fairly general about truth', the following account of how that notion applies to the things he called statements: 'one who makes a statement or assertion makes a true statement if and only if things are as, in making that statement, he states them to be' (Strawson 1970: 180). That does, indeed, seem to be another nice gloss on truth as it applies to statements, and it extends to a cognate account of truth for individual utterances: someone who produces a declarative utterance thereby speaks truly if and only if things are as he thereby says they are. Strawson would have found ascriptions of truth to utterances more than faintly barbarous. However, if we continue to permit ourselves the use of '*u* is true' and '*u* says that *P*' as abbreviations of 'in producing *u*, its speaker speaks truly' and 'in producing *u*, the speaker thereby expresses the thought that *P*', we reach the following:

(*T*) *u* is true if and only if things are as *u* says they are

(*F*) *u* is false if and only if things are not as *u* says they are.

(*T*) and (*F*) are recognizably a version of the Aristotelian idea from which we started and it is this version that I shall develop.

How do (*T*) and (*F*) differ from Ramsey's (T_R) and (F_R)? Well, let us look back at (T_R). According to (T_R), the claim that *u* is true is equivalent to an existentially quantified conjunction, whose first conjunct is '*u* says that *P*'. There is a natural sense, then, in which (T_R) counts '*u* says something' as part of the content of '*u* is true'. (I shall spell out this sense shortly.) Similarly, if (F_R) is correct, then part of the content of '*u* is false' is that *u* says something. By Ramsey's lights, then, a speaker who asserts '*u* is true' (or '*u* is false') thereby asserts that *u* says something, i.e. that *u*

has content. Strawson's formulae, by contrast, offer a subtly different account of this matter. An assertion of 'Things are as *u* says they are', or of 'Things are not as *u* says they are', is not naturally heard as *asserting* that there is a way *u* says things are. Rather, it will be understood as *presupposing* as much, i.e. presupposing that *u* says something. A speaker who asserts 'Things are as that utterance says they are' takes it for granted—and expects his audience to take it for granted—that the utterance in question says something, or has content.

The usual tests confirm that *u*'s having content is a presupposition of Strawson's accounts of its truth and its falsity. 'If things are as John's statement says they are, then Mary's testimony is discredited' implies that there is a way John's statement says things are. So does 'Things are not as John's statement says they are'. Moreover, it seems clear that the failure of this presupposition is 'catastrophic'. That is, if *u* lacks content then the ascription of Strawsonian truth (or falsity) to *u* will itself lack content. According to Strawson, '*u* is true' means 'Things are as *u* says they are'; and if there is no way *u* says things are, there is nothing left in the claim that things are as *u* says they are. Not all failures of presupposition are catastrophic in this sense, but it is pretty clear that this one is.

The difference between Strawson's account of truth and Ramsey's runs parallel to that between Strawson's theory of singular definite descriptions and Russell's. According to Russell (1905), part of what is asserted when someone asserts 'The present king of France is wise' is that there is one and only one king of France. That is why Russell classifies the whole assertion as straightforwardly false. According to Strawson (1950), by contrast, that there is a unique king of France is no part of what is asserted by 'The present king of France is wise'. Rather, it is a presupposition of that assertion. In this particular example, indeed, Strawson held that presupposition failure is catastrophic: since there is no king of France, 'The present king of France is wise' lacks a truth-value altogether.

Strawson's theory of descriptions, however, is questionable in a way that his account of truth is not. It was never Strawson's position that *any* statement containing an empty singular definite description lacks a truth-value. In 'On Referring', he claims this only for 'uniquely referring uses' (or attempted uses) of definite descriptions, and he gives 'Napoleon was the greatest French soldier' as an example of a different sort of use (Strawson 1950: 1). This raises the question of when a statement containing an empty description lacks a truth-value. Strawson comes closest

to giving an answer in 'Identifying Reference and Truth-Values' (1964), where he claims that, in a context in which the topic is *what notable visitors the Exhibition has had*, the statement 'The Exhibition was visited yesterday by the king of France' is false, not truth-value-less, for the 'guilty referring expression' is 'absorbed into the predicate' (Strawson 1964: 89–91). However, he gives no criterion for when such absorption occurs. Without one, his theory of descriptions is seriously incomplete, but his account of truth does not share this deficiency: as we have seen, it is plausible to hold that *whenever u* lacks a content, so does the ascription 'Things are as *u* says they are'.

Despite this, the analogy between Strawson's account of truth and his theory of descriptions will prove to be suggestive as we develop the account, and a first step in that development is to locate more precisely the difference between Ramsey's and Strawson's accounts of truth. Let us say that *A entails B* when it is both the case that (i) *B* follows logically from *A* and (ii) the negation of *A* follows logically from the negation of *B*. Assuming that the inferential behaviour of sentential quantifiers parallels that of their first-order analogues, '$\exists P$(*u* says that *P*)' follows from '$\exists P$(*u* says that $P \wedge P$)'. Moreover, assuming that negation behaves classically, '$\neg\exists P$(*u* says that $P \wedge P$)' follows from '$\neg\exists P$(*u* says that *P*)'. Given these assumption, then, we can say this: on Ramsey's theory of truth, the truth of a declarative utterance *entails* that it has content. It is in this sense (no doubt attenuated) that the claim that *u* has content is part of what is said by '*u* is true'.

On Strawson's account of truth, by contrast, there is no such entailment. Condition (ii) for *A* to entail *B* is that the negation of *A* follows logically from the negation of *B*, and in the present case this condition is not met. For suppose that *u* has no content. From that supposition, it does not follow that things are *not* as *u* says they are. To the contrary, the claim that things are *not* as *u* says they are presupposes that there is a way *u* says things are, just as does the corresponding affirmative claim. We have, then, a clear difference between Strawson's account of truth and Ramsey's. The difference is rather subtle[3] but, as I hope to show, it is important and far-reaching.

[3] It is subtle enough that I overlooked it for years; but then Strawson himself downplayed it. In a late backward glance at his writings on truth, he wrote that by 1950 he was 'already convinced that Ramsey had got the matter essentially right' (Strawson 1998: 8).

6.3 Strawson's account formalized

On Strawson's theory, '*u* is true' presupposes that *u* has content. I have yet to give a positive account of what 'presupposes' means. Many philosophers now follow Stalnaker (1974) in attaching a pragmatic sense to the term. On this conception, the claim that '*u* is true' presupposes that *u* has content amounts to the claim that a speaker who asserts '*u* is true' will ordinarily be understood as taking it for granted that *u* has content. This is part of what I mean, but not the whole of it, for a Stalnakerian presupposition may fail without that failure being catastrophic. In elaborating Strawson's account, we need to articulate the sense in which a statement *A* presupposes *B* in such a way that *B*'s failing to be true deprives *A* of truth-evaluable content.

To specify this relationship, we shall need to work in a logic which makes provision for statements that are neither true nor false, i.e. that are 'undefined'. For definiteness, and to avoid unnecessary duplication of labour, I shall work in an extension of the system of 'simple partial logic' set out by Stephen Blamey in the Second Edition of the *Handbook of Philosophical Logic* (Blamey 2002). The beauty of Blamey's presentation, which owes much to prior work by Dana Scott (especially Scott 1975), is that the semantic clauses for the familiar connectives and first-order quantifiers are precisely the familiar classical clauses. The only changes from the classical semantics are these. First, it is not assumed that every atomic formula is either true or false, or that every *n*-place predicate is true or false of an *n*-tuple of objects; an atomic formula or predication may be undefined. Second, whenever the recursive semantic clauses fail to determine a complex formula as being either true or false, it too is classified as undefined. Thus the semantic clause for conjunction says:

$A \wedge B$	is true	if *A* is true and *B* is true
	is false	if *A* is false or *B* is false
	is undefined	otherwise.[4]

[4] If we think of 'undefined' as a third alethic status, Blamey's clauses for the standard connectives and quantifiers are equivalent to the so-called 'Strong Kleene' semantics for them.

Blamey's 'Fregean' clause for the first-order universal quantifier assumes 'that [the relevant] language has—or can be extended so as to have—a name $\bar{a}$ for each object a in [the domain of quantification] D':[5]

$\forall x\phi(x)$	is true	if $\phi(\bar{a})$ is true for every a in D
	is false	if $\phi(\bar{a})$ is false for some a in D
	is undefined	otherwise

(Blamey 2002: 263)

Simple partial logic has only first-order quantifiers. We shall work in an extended system, to be called simply 'partial logic', which also contains sentential quantifiers under Prior's interpretation. We may adapt Blamey's method in order to give semantic clauses for these quantifiers. Let us assume that our language has—or can be extended so as to have—a sentence $\bar{w}$ corresponding to each entity w in the domain, W, over which the sentential quantifiers range. (Since each such w is a way things might be conceived to be, $\bar{w}$ will be 'Things are in way w'.) We may then give the semantic clauses for those quantifiers as follows:

$\forall P\Phi(P)$	is true	if $\Phi(\bar{w})$ is true for every w in W
	is false	if $\Phi(\bar{w})$ is false for some w in W
	is undefined	otherwise
$\exists P\Phi(P)$	is true	if $\Phi(\bar{w})$ is true for some w in W
	is false	if $\Phi(\bar{w})$ is false for every w in W
	is undefined	otherwise.

These 'Fregean' specifications of truth-conditions do not treat the sentential quantifiers as substitutional: they advert to a domain of extra-linguistic entities (namely, ways things might be), not to a substitution class of expressions. A fuller treatment would describe this domain's structure but this is not needed for present purposes.[6]

I shall follow Blamey in formalizing partial logic as a sequent calculus with a double-barrelled condition for correctness: the sequent $X \Rightarrow Y$ is correct just when (a) some member of Y is true when all the members of X are true, and (b) some member of X is false when all the members of Y

[5] For the distinction between 'Fregean' and 'Tarskian' accounts of the quantifiers which Blamey invokes, see Evans 1977, especially Section II.

[6] For the description I prefer, which allows ways things might be (or might be thought to be) not to be fully determinate, see Rumfitt 2015, especially chapter 6.

are false. A correct sequent, then, both transmits truth forwards along the arrow and transmits falsity backwards. The condition for (multiple-conclusion) logical consequence is similarly double-barrelled. $X \models Y$ if and only if (a) some member of Y is true in every interpretation in which all the members of X are true, and (b) some member of X is false in every interpretation in which all the members of Y are false (Blamey 2002: 265, 332). This definition has the considerable merit of ensuring that $A \dashv\models B$ holds if and only if A and B are logically equivalent in the sense of having the same status (true, false, or undefined) in every interpretation (Blamey 2002: 266). As Blamey remarks, the pure language of partial logic (without add-ons like the intensional operator '*Say*') 'admits "substitutivity of equivalents": when a subformula is replaced by something [logically] equivalent, then the resulting formula is equivalent to the original one' (Blamey 2002: 327). This remains the case when the set of logical operators is enlarged with the addition of sentential quantifiers.

In addition to the familiar operators, Blamey introduces some that are specific to partial logic. For our purposes, the one that matters is the two-place connective '/' of *transplication*, which is defined as follows (Blamey 2002: 264):

A / B	is true	if A is true and B is true
	is false	if A is true and B is false
	is undefined	otherwise (i.e. if A is not true or B is undefined).

In $\ulcorner A / B \urcorner$, A is called the transplicator and B the transplicand.

It is this connective which enables us to formalize the claims about presupposition that lie at the heart of Strawson's reply to Russell. Let 'Kx' symbolize 'x is a present king of France' and 'Wx' symbolize 'x is wise'; then Strawson's analysis of 'The present king of France is wise' may be formalized as

$$\exists x \forall y (Ky \leftrightarrow x = y) / \forall y (Ky \rightarrow Wy)$$

or, in an equivalent existential form, as

$$\exists x \forall y (Ky \leftrightarrow x = y) / \exists y (Ky \wedge Wy).$$

These are equivalent for, in a circumstance where there is no unique present king of France, each of these formulae will be undefined whereas, in a circumstance where there is one king, each formula will be true or

false according as that individual is, or is not, wise. We may, then, read ⌜A/B⌝ as meaning 'B, strongly presupposing that A', where a strong presupposition is one whose failure is catastrophic in the sense of depriving the whole utterance of truth-value. For ⌜A/B⌝ is undefined whenever A is not true.

In using the '/' connective to symbolize Strawson's account of truth, we need to settle what is presupposed by 'Things are as u says they are'. Ramsey's account might lead one to expect the presupposition to be simply '$\exists P$(u says that P)', but further reflection casts doubt on this. It is not enough that there should be *some* way things might be thought to be such that u says things are that way. In order for the presupposition of 'Things are as u says they are' to be fulfilled, there must be such an entity as *the* way u says things are. Only if this requirement is met will there be a single condition that is necessary and sufficient for u to be true.

Quite how to spell out this requirement of uniqueness is a delicate matter. The intuitive idea is that u says one thing when there is a way it says things are and any other way it says things are amounts to the same. The notion of 'amounting to the same' is in acute need of philosophical analysis,[7] but for present purposes we can bracket this problem by introducing an intensional connective '$\equiv$', understood so that '$P \equiv Q$' means 'saying that P amounts to saying that Q'. If 'u says that P' is symbolized as '$Say\ (u,\ P)$', the claim that u says one thing may then be formalized as follows:

$$\exists P \forall Q\, (Say\, (u,\ Q) \leftrightarrow P \equiv Q).$$

On Strawson's account, that u says one thing is a presupposition of both 'u is true' and 'u is false'. As remarked earlier, failure of this presupposition will be catastrophic for ascriptions of both truth and falsity: if there is no such thing as the way u says things are, then a speaker who utters 'Things are as u says they are' himself says nothing. We may, then, formalize the Strawsonian accounts of truth and falsity as the following double sequents:

$$(T) \quad T(u) \Leftrightarrow \exists P \forall Q(Say(u, Q) \leftrightarrow P \equiv Q)/\forall Q(Say(u, Q) \rightarrow Q)$$

$$(F) \quad F(u) \Leftrightarrow \exists P \forall Q(Say(u, Q) \leftrightarrow P \equiv Q)/\forall Q(Say(u, Q) \rightarrow \neg Q)$$

[7] For some discussion see Rumfitt 2016.

or, equivalently, in these existential forms:

$$(T) \quad T(u) \Leftrightarrow \exists P \forall Q(Say(u, Q) \leftrightarrow P \equiv Q)/\exists Q(Say(u, Q) \wedge Q)$$

$$(F) \quad F(u) \Leftrightarrow \exists P \forall Q(Say(u, Q) \leftrightarrow P \equiv Q)/\exists Q(Say(u, Q) \wedge \neg Q).$$

The parallel with Strawson's treatment of singular definite descriptions will be evident. I have formalized (*T*) and (*F*) as sequents rather than bi-conditionals, for $\ulcorner A \leftrightarrow B \urcorner$ is undefined when both *A* and *B* are. The addition to partial logic of the sequents (*T*) and (*F*) cannot generate inconsistency, for any such sequent is correct in an interpretation where no utterance *u* says one thing; in such an interpretation, both antecedent and succedent are undefined for all values of *u*.

Strawson's account may seem to give the wrong results when applied to things that are not meaningful utterances: we surely want 'Buckingham Palace is true' to come out straightforwardly false, not as undefined. However, we can achieve this result simply by filtering out those things that are not even meaningful utterances, deeming them all to be not true, and applying Strawson's analysis only to the things that remain. This is similar to the approach Saul Kripke took in his 'Outline of a Theory of Truth' where all entities that are not 'sentences' are assigned to the 'anti-extension' of the truth-predicate, i.e. are classified as not true (see Kripke 1975: 702). I prefer to speak of meaningful utterances rather than sentences, for it seems to me that a nonsensical but grammatical sentence, such as Lewis Carroll's 'All mimsy were the borogoves',[8] is also a case of not being true; however, the filtration can be adjusted to accommodate different views about this. The important point is that Strawson's account gives different verdicts from Ramsey's in cases which are *candidates* for assessment as true or as false, but which fail to say one thing.

6.4 The Liar and the Liar's Revenge

In order to illustrate this, consider what happens when Strawson's account of truth is applied to a paradoxical utterance. Let λ be an

[8] Humpty Dumpty does explain to Alice that '"borogove" is a thin, shabby-looking bird with its feathers sticking out all around, something like a live mop', but if Putnam (1973) and Wiggins (1993) are right about terms for biological species, this explanation does not specify the word's sense. On the Putnam-Wiggins view, the sense of a species term is 'object-involving': to be a borogove is to belong to the same species as a given exemplar. In the present case, there are no exemplars.

utterance of 'This utterance is not true'. It seems clear that *if* λ says one thing, then the thing it says is that λ is not true. But then the hypothesis that λ says one thing generates a contradiction in a familiar way. For suppose λ is true. In that case, things must be as it says they are, so λ is not true. Thus the supposition that λ is true yields a contradiction, so λ is not true. Given that λ says something, though, what it says is that λ is not true. So things *are* as λ says they are, whence λ is true after all. The natural conclusion to draw from this contradiction is that λ does not say one thing.

We may formalize this argument as a proof in partial logic if we assume that the 'not' in 'This utterance is not true' corresponds to the negation operator '¬' of partial logic whose truth- and falsity-table is as follows:

$\neg A$	is true	if A is false
	is false	if A is true
	is undefined	otherwise (i.e. if A is undefined).

This negation operator 'passes through' a transplication in the sense that $\neg(A/B)$ is logically equivalent to $A/\neg B$. By applying this result to the existential form of Strawson's account of truth, we reach the result that

$$\neg T(u) \Leftrightarrow \exists P \forall Q(Say(u, Q) \leftrightarrow P \equiv Q)/\neg\exists Q(Say(u, Q) \wedge Q)$$

and hence

$$\neg T(u) \Leftrightarrow \exists P \forall Q(Say(u, Q) \leftrightarrow P \equiv Q)/\forall Q(Say(u, Q) \rightarrow \neg Q).$$

This last result yields $\neg T(u) \Leftrightarrow F(u)$.

Given that there is no one thing which λ says, should we infer that it says nothing, or that it says more than one thing? Some writers have thought that the key to resolving the Liar is to recognize that an utterance such as λ expresses more than one thought. Thomas Bradwardine argued for this in the fourteenth century and Stephen Read has recently tried to revive his solution (see Read 2009). According to Bradwardine, every saying, as well as expressing its literal content, expresses its own truth. Thus λ expresses both the thought that λ is not true (its literal content) and the distinct—indeed, the contradictory—thought that λ is true. This offers an interesting way out of the paradox. In the spirit of the universal form of (*T*), Bradwardine's requirement for an utterance to be true was that *anything* it says should be the case. So, while we may conclude that λ

is not true, we cannot infer from that that it is true. For present purposes, however, we need not adjudicate between this view and one on which λ expresses no thought at all. It will be enough to have established that there is no one thing which λ says.

The reasoning that establishes this conclusion goes through with only minor modifications on Ramsey's account of truth. A follower of Ramsey can and should accept the premiss of our derivation—namely, that if λ says one thing, that one thing is tantamount to the claim that λ is not true. Since the hypothesis up for refutation is precisely that the Strawsonian presupposition of λ's truth is fulfilled, when deducing consequences from that hypothesis we can exclude the possibility—that of presupposition failure—in which Ramsey's account of truth deviates from Strawson's. Many Ramseyans have, indeed, embraced the conclusion. Williamson, for example, holds that λ says nothing at all. On his view, 'the semantic paradoxes are transformed into sound arguments for what can [and cannot] say what in what contexts' (1998: 15).

There is, however, a further twist to the paradox and it is here that the two accounts of truth diverge. Let us suppose, with Williamson, that λ says nothing. In the classical logic that he takes for granted, it follows that λ says nothing that is the case. By contraposing (T_R), we may infer if λ says nothing that is the case, then it is not true. It follows that λ is not true. But now we are again in a bind, the so-called 'Liar's Revenge'. For, by classically valid logical steps, Williamson can be led to make a statement—namely, 'λ is not true'—which by the lights of his own theory is without content.

I shall soon consider whether these steps really are valid. The first point to note, though, is that the Revenge problem does not arise on Strawson's account of truth. The argument given above shows that λ says no one thing. On Strawson's account, however, that λ says one thing is a presupposition of 'λ is not true', just as it is a presupposition of 'λ is true'. Since the presupposition fails, 'λ is not true' lacks a truth-value, just as λ does. There is no sound deduction from 'λ says no one thing' to 'λ is not true'.

Some people will protest that this exposes an expressive inadequacy in the language of partial logic. Given that λ says nothing, they still want to assert 'λ is not true' and they think there ought to be an 'external' negation operator which enables them to assert this. But is there? Empirical investigations into English and other tongues suggest that

such an operator is not found naturally (see Tappenden 1999, which draws on Horn 1989). Moreover, a moral of the Liar is that trying to introduce one artificially will generate contradictions. So the desire to assert that λ is not true should not—indeed, cannot—be gratified. On this analysis, the Liar reveals a limitation on what it is possible to express, but the limitation is not seriously restrictive. We can, and do, live without external negation.

If an objector insists that there must be some way of saying that λ is not true, we can go halfway to meet him, for there is nothing to stop us from *denying* the truth of λ. On our analysis, λ itself, its negation, and any statement that ascribes truth to it are one and all undefined: they say nothing. We are within our rights, then, to deny any of these utterances, where to deny a formula is to reject it as having some status other than truth. We shall be snared in the toils of paradox only if we postulate a negation operator, '~', for which asserting $\ulcorner{\sim}A\urcorner$ is tantamount to denying $\ulcorner A\urcorner$. We should rest content with our denial and not try to factorize it into the assertion of any species of negation (see further Parsons 1984).

What, though, about the Revenge problem for Ramsey's account of truth? We cannot ignore this, for even if Ramsey misdescribes our ordinary concept of truth, there is nothing to stop us from defining '*u* is Ramsey-true' to mean '$\exists P(Say(u, P) \wedge P)$' and then taking λ to be an utterance of 'λ is not Ramsey-true'. Or, more directly, we may simply take λ to be an utterance of $\neg\exists P(Say(\lambda, P) \wedge P)$. As noted, the problem is that we seem to be able to prove that this λ says nothing—but also prove what it says.

The most popular solution rests on the observation—which is undeniably correct in general—that the domain over which a quantifier ranges may vary with the context of utterance. Contradiction may then be avoided so long as the range of the quantifier '$\exists P$' in the original utterance, λ, does not extend to cover whatever might be said by an inscription of $\neg\exists P(Say(\lambda, P) \wedge P)$ which is part of an analysis of, or commentary upon, λ. (For this approach, see Parsons 1974.) It is, though, hard to see why the range of the original '$\exists P$' *cannot* cover what might be said by such an inscription. We usually take a quantifier's range to be determined by the speaker's intentions, and the producer of λ might make it clear that he does intend the range of its component '$\exists P$' to cover whatever might be expressed by a commentary on, or an analysis of, λ. In the absence of some explanation of why this apparent

possibility is not in fact possible, this contextualist strategy provides no satisfactory resolution of the Liar's Revenge.[9]

It is, then, of some interest that partial logic provides an alternative way out. The Revenge problem turns on the inference from $\neg\exists P Say(\lambda, P)$ to $\neg\exists P(Say(\lambda, P) \wedge P)$, but this inference is not sound in partial logic: it may be that the premiss is true but the conclusion is undefined. Since the logic does validate the inference from $X \Rightarrow Y$ to $\neg Y \Rightarrow \neg X$ (where $\neg X$ is the set of all negations of formulae in X), the sequent $\exists P(Say(\lambda, P) \wedge P) \Rightarrow \exists P Say(\lambda, P)$ must also be incorrect: it may fail to transmit falsity back from the succedent to the antecedent. This in turn requires some restriction on the usual sequent rules for the existential sentential quantifier. Given the normal rules for conjunction, we cannot have

($\exists P1$) From $X, \exists P\Phi \Rightarrow Y$, infer $X, \Phi(A/P) \Rightarrow Y$

and

($\exists P2$) From $X, \Phi(A/P) \Rightarrow Y$, infer $X, \exists P\Phi \Rightarrow Y$,
provided that 'P' is not free in the second sequent

on pain of being able to derive the paradoxical sequent $\exists P(Say(\lambda, P) \wedge P) \Rightarrow \exists P Say(\lambda, P)$. Our framework, though, enables us to state the needed restrictions on the quantifier rules. Remark first that, for any sentence B, '$X, B \Rightarrow \neg B, Y$' amounts to saying 'Given B, the sequent $X \Rightarrow Y$ is correct'. Remark, second, that '$A \equiv A$' is true if and only if A says something. The rules for the existential quantifier are then

($\exists P1'$) From $X, \exists P\Phi \Rightarrow Y$, infer $X, \Phi(A/P), A \equiv A \Rightarrow \neg(A \equiv A), Y$

and

($\exists P2'$) From $X, \Phi(A/P), A \equiv A \Rightarrow \neg(A \equiv A), Y$, infer $X, \exists P\Phi \Rightarrow Y$,
provided that 'P' is not free in the second sequent.[10]

[9] On the contextualist strategy, see further Rumfitt 2017: §4.

[10] We need corresponding restrictions on the rules for the universal sentential quantifiers, viz.

($\forall P1'$) From $X, A \equiv A \Rightarrow \neg(A \equiv A), \Phi(A/P), Y$, infer $X \Rightarrow \forall P\Phi, Y$,
provided that P is not free in the second sequent

and

($\forall P2'$) From $X \Rightarrow \forall P\Phi, Y$ infer $X, A \equiv A \Rightarrow \neg(A \equiv A), \Phi(A/P), Y$.

These revised rules block the derivation of the paradoxical sequent; the derivation requires the premiss '$\neg\exists P(Say(\lambda, P) \wedge P) \equiv \neg\exists P(Say(\lambda, P) \wedge P)$', which we have not got. In partial logic, then, the Revenge argument does not go through.

Is it, though, worth deviating from classical logic in order to avoid the Liar's Revenge? In an important sense, the solution involves no such deviation. Using the sentential quantifiers recommended for this purpose at the end of §1, we may still assert Excluded Middle in the form $\forall P(P \vee \neg P)$: however things may be conceived to be, either things are thus or they are not thus. To be sure, not every instance of the schema $\ulcorner A \vee \neg A \urcorner$ is true, but that is because some well-formed replacements for A fail to say that things are thus-and-so. The revised inference rules for the quantifiers are needed in order to allow for cases where we have such a failure, but these cases are no more counterexamples to classical logical laws than Pegasus is a counterexample to the biological law that horses lack wings. In this, I am at one with Kripke: 'conventions for handling sentences that do not express propositions are not in any philosophically interesting sense "changes in logic" . . . All our considerations can be formalized in a classical metalanguage' (Kripke 1975: 700–1, n. 18).

As Kripke recognized, the present solution depends upon distinguishing between the object language and the metalanguage. The theory classifies λ as undefined and hence as not true, 'but we are precluded from saying this in the object language by our interpretation of negation and the truth predicate. If we think of the minimal fixed point, say under the Kleene valuation, as giving a model of natural language, [the claim] that a Liar sentence is not true must be thought of as associated with some later stage in the development of natural language, one in which speakers reflect on the generation process leading to the minimal fixed point. It is not itself a part of that process. The necessity to ascend to a metalanguage may be one of the weaknesses of the present theory' (1975: 714). The matter deserves far closer attention than I can give it here, but I doubt if it is a weakness. Hierarchies of languages have a bad name because, on Tarski's account of them (1935, 1944), no sentence which ascribes truth to any sentence of L can itself belong to L. The present approach, like Kripke's, rejects that draconian restriction. Instead, it adopts the quite different idea that an account of how a given language is able to contain its own truth-predicate may need to be expressed in a different language. If we are forced to contemplate an indefinitely extensible hierarchy of languages, that will

reflect the open-endedness of philosophical explanation, not the impossibility of ascribing, in *L*, truth to the utterances of *L*.[11]

6.5 Axiomatic theories of truth

Since 1975, when Kripke published his paper, philosophical and logical approaches to the topic of truth have, regrettably, diverged. Philosophers strive, as they always have, to say in general terms what it is for the various bearers of truth to be true, but many logicians now disdain such elucidations in favour of axioms which regulate the behaviour of a truth-predicate as it applies to the formulae of particular formalized languages. A typical axiom of this kind says that a conjunctive formula is true if and only if both its conjuncts are true. Philosophers are unlikely to contest the truth of this, but they may reasonably object to its being taken as axiomatic. 'Surely', they may say, 'this so-called axiom ought to be grounded in a general elucidation of truth—something that can account for its application to truth-bearers other than well-formed formulae—together with more particular principles about the meaning of conjunctions'. In reply, though, logicians are apt to protest that the general elucidations philosophers have come up with are too woolly to be useful. Thus Leon Horsten, a leading proponent of the axiomatic approach, complains that 'all substantial [philosophical] theories suffer from a lack of formal precision' (Horsten 2011: 13). I shall defend the present account against this charge by showing how Horsten's preferred principles about truth and falsity are in fact theorems of a theory whose axioms are versions of (*T*) and (*F*) along with highly plausible axioms about what formulae say.

I say *versions* of (*T*) and (*F*), for we have to be careful about the categories to which truth and falsity apply. (*T*) and (*F*) describe the application of these notions to individual declarative utterances. Like most contemporary logicians, though, Horsten follows Tarski in applying them to type sentences. Not all type sentences are sensibly classified as true or false

[11] There are other issues that need to be addressed in defending the recommended treatment of the Liar. One is this. According to Blamey's semantics, the disjunction of 'Snow is white' with λ is true. Since it is true, the disjunction surely expresses a thought—but which? The thought that snow is white, or something more complex? An important but neglected distinction of Dummett's helps defend the second alternative: the fact that λ has no 'assertoric content' does not entail that it lacks 'ingredient sense' (see Dummett 1981: 446–7). However, the task of applying that distinction to the present case must await another paper.

simpliciter; it makes no sense so to classify the English type sentence 'I am ill'. However, after spelling out the appropriate restriction, we can extend Strawson's account of truth and falsity so that it covers these Tarski-style applications of the notions. Let us call a type sentence *eternal* if all utterances of it which express a thought at all express the same thought. Let an eternal sentence *say* that *P* (written: *Says*(⌜*A*⌝, *P*)—note the terminal 's' in '*Says*') if and only if any utterance of it expresses the thought that *P*. (For reasons that will emerge, it helps to interpret the notion of *saying* so strictly that, while the German type sentence 'Schnee ist weiß und Gold ist gelb' says that snow is white and gold is yellow, it does not count as saying that gold is yellow and snow is white.) Let us call an eternal type sentence *contentful* when it says something, i.e. when $\exists P\, Says(\ulcorner A \urcorner, P)$. A contentful eternal type sentence will express the same thought in all contexts. Let '*Tr*' and '*Fls*' symbolize truth and falsity as they apply to eternal type sentences. Then we may lay down the equivalences (*Tr*) and (*Fls*) as Strawsonian accounts of this sort of application of truth and falsity:

(*Tr*) $Tr(\ulcorner A \urcorner) \Leftrightarrow \exists P Says(\ulcorner A \urcorner, P)/\forall Q(Says(\ulcorner A \urcorner, Q) \rightarrow Q)$

(*Fls*) $Fls(\ulcorner A \urcorner) \Leftrightarrow \exists P Says(\ulcorner A \urcorner, P)/\forall Q(Says(\ulcorner A \urcorner, Q) \rightarrow \neg Q)$.

(*Tr*) asserts the equivalence of 'Eternal type sentence *A* is true' with '*A* is contentful / However *A* says things are, they are thus' and (*Fls*) asserts the equivalence of 'Eternal type sentence *A* is false' with '*A* is contentful / However *A* says things are, they are not thus'.

Since a contentful eternal type sentence expresses precisely one thought, we also have the alternative existential forms:

(*Tr*) $Tr(\ulcorner A \urcorner) \Leftrightarrow \exists P Says(\ulcorner A \urcorner, P)/\exists Q(Says(\ulcorner A \urcorner, Q) \wedge Q)$

(*Fls*) $Fls(\ulcorner A \urcorner) \Leftrightarrow \exists P Says(\ulcorner A \urcorner, P)/\exists Q(Says(\ulcorner A \urcorner, Q) \wedge \neg Q)$.

The existential form of (*Tr*) asserts the equivalence of 'Eternal type sentence *A* is true' with '*A* is contentful / *A* says that things are somehow and they are thus'.

As with (*T*) and (*F*), the theory that results from adding the sequents (*Tr*) and (*Fls*) to partial logic is formally consistent. Any such sequent is vacuously correct in an interpretation where no eternal sentence in the relevant language says anything (cf. Williamson 1998: 14).

It is these equivalences, (*Tr*) and (*Fls*), which combine with plausible theses about which thoughts eternal type sentences of various forms express to yield the principles about truth and falsity which Horsten

takes as axiomatic. His preferred principles are those of the system called **PKF** ('Partial Kripke Feferman'), which explicitly axiomatizes the fixed-point account of truth whose models Kripke characterized in 'Outline of a Theory of Truth'. The background logic of **PKF** is simple partial logic, and its primitive connectives are '¬', '∧', '∨', and the first-order quantifiers. Its axioms include all those of first-order Peano Arithmetic, so the theory can describe the syntax of its own formulae. The system also contains a number of 'truth rules' whose soundness amounts to the correctness of the sequents **PKF1** to **PKF7**:

PKF1 $Tr(\ulcorner t_1 = t_2 \urcorner) \Leftrightarrow$ the denotation of t_1 = the denotation of t_2
PKF2 $Tr(\ulcorner A \wedge B \urcorner) \Leftrightarrow Tr(\ulcorner A \urcorner) \wedge Tr(\ulcorner B \urcorner)$
PKF3 $Tr(\ulcorner A \vee B \urcorner) \Leftrightarrow Tr(\ulcorner A \urcorner) \vee Tr(\ulcorner B \urcorner)$
PKF4 $Tr(\ulcorner \forall x \phi(x) \urcorner) \Leftrightarrow \forall x Tr(\ulcorner \phi(x) \urcorner)$
PKF5 $Tr(\ulcorner \exists x \phi(x) \urcorner) \Leftrightarrow \exists x Tr(\ulcorner \phi(x) \urcorner)$
PKF6 $Tr(\ulcorner Tr(A) \urcorner) \Leftrightarrow Tr(\ulcorner A \urcorner)$
PKF7 $Tr(\ulcorner \neg A \urcorner) \Leftrightarrow \neg Tr(\ulcorner A \urcorner)$

(see Horsten 2011: 132–5, and Halbach and Horsten 2006: 692–3). No separate rules are given for falsehood; rather, $Fls(\ulcorner A \urcorner)$ is defined to be equivalent to $Tr(\ulcorner \neg A \urcorner)$. In §4, we already observed that on Strawson's theory, the falsehood of a formula is demonstrably equivalent to the truth of its negation.

Each of the sequents **PKF1–7** may be derived from (*Tr*) in tandem with appropriate additional sequents which specify what type sentences of various forms say. I have space here only to sketch derivations of **PFK2** and **PKF7**. I shall use $Cont(\ulcorner A \urcorner)$ as an abbreviation for $\exists P Says(\ulcorner A \urcorner, P)$.

In deriving **PKF2**, the truth-rule for conjunctions, we apply an equivalent of the conjunction of two transplications:

$$(A_1/B_1) \wedge (A_2/B_2) \Leftrightarrow ((A_1 \wedge A_2) \vee (A_1 \wedge \neg B_1) \vee (A_2 \wedge \neg B_2))/B_1 \wedge B_2)$$

(see Blamey 2002: 328).[12] By (*Tr*),

[12] Empirical linguists have sought 'projection rules' which determine the presuppositions of complex sentences, given the presuppositions of their parts (the seminal paper is Karttunen 1973). The transplicator of the right-hand formula in the sequent displayed is equivalent in partial logic to the result of applying the 'symmetric' projection rule for conjunction that some empirical linguists recommend (see e.g. Rothschild 2008). The same correspondence applies in the case of disjunctions. This encourages the hope that the logical approach to presupposition followed here is not completely divorced from the pragmatic conception of the topic that now prevails in linguistics.

$Tr(\ulcorner A \urcorner) \wedge Tr(\ulcorner B \urcorner) \Leftrightarrow$
$(Cont(\ulcorner A \urcorner)/\forall P(Says(\ulcorner A \urcorner, P) \rightarrow P) \wedge$
$(Cont(\ulcorner B \urcorner)/\forall Q(Says(\ulcorner B \urcorner, Q) \rightarrow Q)).$

By the equivalence just cited, this yields

$Tr(\ulcorner A \urcorner) \wedge Tr(\ulcorner B \urcorner) \Leftrightarrow$
$(Cont(\ulcorner A \urcorner) \wedge Cont(\ulcorner B \urcorner)) \vee (Cont(\ulcorner A \urcorner) \wedge \neg\forall P(Says(\ulcorner A \urcorner, P) \rightarrow P)) \vee$
$(Cont(\ulcorner B \urcorner) \wedge \neg\forall Q(Says(\ulcorner B \urcorner, Q) \rightarrow Q))/$
$\forall P(Says(\ulcorner A \urcorner, P) \rightarrow P) \wedge \forall Q(Says(\ulcorner A \urcorner, Q) \rightarrow Q).$

The transplicator of the right-hand formula in the last sequent may look a mess. However, the conjunction $\ulcorner A \wedge B \urcorner$ will be contentful if and only if either $\ulcorner A \urcorner$ and $\ulcorner B \urcorner$ are both true, or $\ulcorner A \urcorner$ is false, or $\ulcorner B \urcorner$ is false.[13] Each of these three possibilities corresponds to a disjunct in the transplicator, which accordingly reduces to $Cont(\ulcorner A \wedge B \urcorner)$. Thus we reach

$Tr(\ulcorner A \urcorner) \wedge Tr(\ulcorner B \urcorner) \Leftrightarrow$
$Cont(\ulcorner A \wedge B \urcorner)/\forall P(Says(\ulcorner A \urcorner, P) \rightarrow P) \wedge \forall Q(Says(\ulcorner A \urcorner, Q) \rightarrow Q).$

We also postulate, as a principle about what conjunctions express,

(*Conj*) $Says(\ulcorner A \wedge B \urcorner, P \wedge Q) \Leftrightarrow Says(\ulcorner A \urcorner, P) \wedge Says(\ulcorner B \urcorner, Q).$

Given our strict reading of '*Says*', this principle is highly plausible, and it yields

$\forall R(Says(\ulcorner A \wedge B \urcorner, R) \rightarrow R) \Leftrightarrow$
$\forall P(Says(A, P) \rightarrow P) \wedge \forall Q(Says(B, Q) \rightarrow Q).$

Putting these results together we reach

$Tr(\ulcorner A \urcorner) \wedge Tr(\ulcorner B \urcorner) \Leftrightarrow Cont(\ulcorner A \wedge B \urcorner)/\forall R(Says(\ulcorner A \wedge B \urcorner, R) \rightarrow R)$

which by (*Tr*) yields

$Tr(\ulcorner A \urcorner) \wedge Tr(\ulcorner B \urcorner) \Leftrightarrow Tr(\ulcorner A \wedge B \urcorner).$

This is **PKF2**, as required.

[13] Here it is vital that '$\wedge$' is interpreted according to the Strong Kleene tables (cf. n. 4 above). If it were interpreted according to the Weak Kleene tables, then $Cont(\ulcorner A \wedge B \urcorner)$ would be equivalent to $Cont(\ulcorner A \urcorner) \wedge Cont(\ulcorner B \urcorner)$. On that alternative reading, however, $(A_1/B_1) \wedge (A_2/B_2)$ is equivalent to $(A_1 \wedge A_2/B_1 \wedge B_2)$, so we still get the equivalence between $Tr(\ulcorner A \wedge B \urcorner)$ and $Tr(\ulcorner A \urcorner) \wedge Tr(\ulcorner B \urcorner)$.

What about **PKF7**, Horsten's truth-rule for negations? A highly plausible principle about what negated formulae express is

(*Neg*) $Says(\ulcorner A \urcorner, P) \Leftrightarrow Says(\ulcorner \neg A \urcorner, \neg P)$,

which immediately yields

$\exists P Says(\ulcorner A \urcorner, P) \Leftrightarrow \exists P Says(\ulcorner \neg A \urcorner, P)$,

i.e.

$Cont(\ulcorner A \urcorner) \Leftrightarrow Cont(\ulcorner \neg A \urcorner)$.

Now by (*Tr*) we have

$Tr(\ulcorner \neg A \urcorner) \Leftrightarrow Cont(\ulcorner \neg A \urcorner)/\forall Q(Says(\ulcorner \neg A \urcorner, Q) \rightarrow Q)$,

so we reach

$Tr(\ulcorner \neg A \urcorner) \Leftrightarrow Cont(\ulcorner A \urcorner)/\forall Q(Says(\ulcorner \neg A \urcorner, Q) \rightarrow Q)$.

Since any formula is equivalent in partial logic to its double negation, (*Neg*) also yields

$Says(\ulcorner \neg A \urcorner, P) \Leftrightarrow Says(\ulcorner A \urcorner, \neg P)$

so we may infer

$Tr(\ulcorner \neg A \urcorner) \Leftrightarrow Cont(\ulcorner A \urcorner)/\forall Q(Says(\ulcorner A \urcorner, \neg Q) \rightarrow Q)$.

Again given the equivalence of a formula and its double negation, the transplicand of this last formula is equivalent to $\forall Q(Says(\ulcorner A \urcorner, Q) \rightarrow \neg Q)$. Thus we reach

$Tr(\ulcorner \neg A \urcorner) \Leftrightarrow Cont(\ulcorner A \urcorner)/\forall Q(Says(\ulcorner A \urcorner, Q) \rightarrow \neg Q)$.

and hence

$Tr(\ulcorner \neg A \urcorner) \Leftrightarrow Cont(\ulcorner A \urcorner)/\neg \exists Q(Says(\ulcorner A \urcorner, Q) \wedge Q)$.

Now $A/\neg B \Leftrightarrow \neg(A/B)$, so the last sequent yields

$Tr(\ulcorner \neg A \urcorner) \Leftrightarrow \neg(Cont(\ulcorner A \urcorner)/\exists Q(Says(\ulcorner A \urcorner, Q) \wedge Q)$.

The right-hand formula in the last sequent is the negation of the existential form of $Tr(\ulcorner A \urcorner)$. So we reach

$Tr(\ulcorner \neg A \urcorner) \Leftrightarrow \neg Tr(\ulcorner A \urcorner)$,

which is precisely **PKF7**.

The derivation of **PKF7** is especially noteworthy, for many philosophers will reject this sequent's correctness. Let Λ be the paradoxical type sentence $\neg Tr(\ulcorner \Lambda \urcorner)$. Ramseyans will accept that Λ does not express a thought. They will infer from that, however, that it does not express a true thought; that is, they will infer $\neg Tr(\ulcorner \Lambda \urcorner)$. They will not, however, accept $Tr(\ulcorner \neg \Lambda \urcorner)$, so they are committed to rejecting **PKF7**. Horsten and other adherents of **PKF** may be pressed, then, to say why they think **PKF7** is correct. It will not be enough for them to point out that **PKF** is a consistent theory of truth, for there are many of those. What they need is a philosophical elucidation of truth to set against the Ramseyan account with which **PKF7** is incompatible. Strawson's account provides this. For these reasons, the philosopher's quest for such an elucidation cannot, in the end, be avoided. Moreover, as we have seen, that quest may be pursued without any compromise in logical rigour.

The derivation, moreover, illuminates Kripke's theory of truth. Kripke originally expounded that theory by describing its intended models. **PKF** was then presented as a system of axioms which precisely characterizes those models. Our derivation shows that the axioms which achieve this are not artificial, but are consequences of (*Tr*) and (*Fls*)—themselves applications of attractive general elucidations of truth and falsity to eternal type sentences—along with principles about what formulae say. Together, these connections suggest that the notions of truth and falsity that Kripke described play a central role among our semantic concepts.

6.6 Intensional paradox

The derivations sketched in the previous section rely on compositional principles about what type sentences of various forms say. Those principles presume that some simple sentences say things, but what principles regulate this? While no inconsistency arises from adding the sequents (*Tr*) and (*Fls*) to partial logic, we shall fall into inconsistency if we also add wholesale disquotational principles about what sentences say. In particular, we shall get a contradiction if we add $Says(\ulcorner \Lambda \urcorner, \neg Tr(\ulcorner \Lambda \urcorner))$. One residual problem posed by the Liar is to say as much as possible about which sentences say what.

Can we assert disquotational principles for sentences free of semantical vocabulary and propositional attitude verbs? Can we at least assert—call this (*SW*)—that any English speaker who utters the sentence 'Snow is white' thereby says that snow is white? Prior had an argument

which, he thought, refuted even (*SW*) (Prior 1961: 29). Suppose Tarski utters 'Snow is white' immediately after Prior has said 'If, immediately after this, Tarski says something that is the case, then what I am now saying is not the case'. Since snow is white, we reach a contradiction if we suppose both (1) that Tarski says that snow is white, and (2) that Prior says that if Tarski immediately says something that is the case, then what Prior says is not the case. According to Prior, the only possible basis for holding that one of these utterances, rather than the other, succeeds in expressing a thought is 'who got his say in first' (1961: 21). Since Prior spoke first, he made it impossible for Tarski to say anything that is the case, so when Tarski uttered 'Snow is white' he could not thereby say that snow is white. The most we can say about that sentence is that an English speaker who utters it says that snow is white, *provided* he says something by it. If it works at all, this argument could be applied to refute *any* principle to the effect that an utterance of a type sentence *S* always says that *P*. It would thereby scupper systematic semantic theorizing.

I do not think that the argument does work, though: it accords quite unreasonable importance to temporal priority. A more sensible view takes account of whether the relevant sentences involve semantic predicates or propositional attitude verbs. As per (*SW*), when Tarski utters the simple sentence 'Snow is white' he succeeds in saying that snow is white, and the fact that he expresses a truth shows it was Prior who failed to say anything by his utterance. On this view, whether Prior's utterance has a content will depend on whether Tarski subsequently utters a truth or a falsehood, so no judgement on whether Prior's utterance has content is possible before Tarski speaks. This consequence, however, is not implausible. Prior's words advert to Tarski's forthcoming utterance and, given what they mean, we should accept that whether they express a thought depends on the truth-value of that later utterance. To be sure, if the words 'immediately after this' in Prior's scuppering formula were replaced by 'ten years from today', we should have to wait that long to find out whether Prior had said anything. This consequence is certainly odd; then again, the scuppering formula is a very odd thing to say. Most will be prepared to swallow the consequence in order to preserve (*SW*) and, with it, the possibility of a systematic semantic theory.

What further general principles can we lay down about which sentences say what? We may approach this question by analysing another of Prior's intensional paradoxes, his version of the Epimenides (Prior 1958, 1961).

In Prior's telling, Epimenides the Cretan utters the sentence 'Nothing said by a Cretan is the case'. That utterance cannot be true, of course, from which it appears to follow that something said by a Cretan is the case. Suppose, though, that nothing *else* said by a Cretan is the case. In such a circumstance, the supposition that Epimenides succeeds in saying that nothing said by a Cretan is the case yields a contradiction; so it would appear to be 'impossible for Epimenides the Cretan to say that nothing said by a Cretan is the case; whatever noises he makes, he will not under those circumstances be able to say *that* by them' (Prior 1961: 19). As Prior observes, this result appears to be a simple corollary 'of the obvious truth that *if it is a fact that no fact is asserted by a Cretan, then THIS fact (that no fact is asserted by a Cretan) is not asserted by a Cretan either*' (ibid.). There are corresponding truths for other intensional verbs. Thus, if it is a fact that no Cretan believes any fact, no Cretan believe THIS fact (namely, that no Cretan believes any fact). Again, if it is a fact that no Cretan fears anything that is the case, then no Cretan fears that that no Cretan fears anything that is the case. Prior concluded, albeit 'unwillingly', that 'thinking, fearing, etc., because they are attitudes in which we put ourselves in relation to the real world, must from time to time be oddly blocked by factors in that world, and we must just let Logic teach us where these blockages will be encountered' (Prior 1961: 32).

But how, exactly, is logic supposed to teach us this? Let us use 'δP' to symbolize 'It is said by a Cretan that P'. Then Epimenides' utterance may be formalized as $\forall P(\delta P \rightarrow \neg P)$. Now if Epimenides did succeed in saying what his words would normally be taken as saying, then it would be said by a Cretan that nothing said by a Cretan is the case. That is, we would have $\delta(\forall P(\delta P \rightarrow \neg P))$. From this supposition, however, we appear to be able to construct the following derivation:

1.	$\delta(\forall P(\delta P \rightarrow \neg P))$	Assumption
2.	$\forall P(\delta P \rightarrow \neg P)$	Premiss
3.	$\delta(\forall P(\delta P \rightarrow \neg P)) \rightarrow \neg\forall P(\delta P \rightarrow \neg P)$	(2) ∀-elimination, instantiating 'P' with $\forall P(\delta P \rightarrow \neg P)$
4.	$\neg\forall P(\delta P \rightarrow \neg P)$	(1), (3) *modus ponens*
5.	$\neg\delta(\forall P(\delta P \rightarrow \neg P))$	(2), (4) ¬-introduction, discharging assumption (1).[14]

[14] This derivation is modelled on one given by Prior (1961: 19) but I eschew his Polish notation.

From the premiss that nothing said by a Cretan is the case, then, we have an apparent proof that Epimenides did not succeed in saying what his words would normally be taken as saying. Since nothing is assumed about the meaning of 'δ', we also appear to have proofs of the corresponding results for other propositional attitudes.

This analysis depends, though, upon the fact that I (the producer of the derivation) am not Cretan. For if I were Cretan, the eventual conclusion, (5), would preclude me from stating the premiss at line (2). Prior is clear about this: 'the thing itself—that nothing said by a Cretan is the case—will under those circumstances be true . . . That is, there [is] nothing wrong with *what we suppose the Cretan to say*, but only with *the supposition that* he says it' (Prior 1961: 19). His analysis depends on being able to insert a wedge between what I, the analyst or commentator, can say about the situation (in the indirect speech sense of 'say') and what a Cretan can say. But what happens if I am a Cretan, when no such wedge can be inserted?

Partial logic provides an answer. Note first that the simple derivation (1)–(5) is not valid in that logic: its natural deduction formalization does not have an unrestricted rule of ¬-introduction. Using the sequent calculus formalization preferred in this paper, let 'Δ' abbreviate the formula '$\forall P(\delta P \rightarrow \neg P)$'. Then the rules for the universal sentential quantifier (see n. 10 above) yield:

$$\forall P(\delta P \rightarrow \neg P), (\Delta \equiv \Delta) \Rightarrow \neg(\Delta \equiv \Delta), \neg\delta(\forall P(\delta P \rightarrow \neg P))$$

i.e.

$$\forall P(\delta P \rightarrow \neg P), (\Delta \equiv \Delta) \Rightarrow \neg(\Delta \equiv \Delta), \neg\delta\Delta.$$

Given the truth of $\forall P(\delta P \rightarrow \neg P)$ as a premiss, however, we cannot always infer that δΔ is false (as Prior supposes we can). δΔ will, indeed, be false in the case where $\Delta \equiv \Delta$ is true. However, δΔ will be undefined when $\Delta \equiv \Delta$ is not true. These cases corresponds to the two scenarios considered above. Given that I am not a Cretan, $\Delta \equiv \Delta$ is true, so we can infer ¬δΔ. If I am a Cretan, though, $\Delta \equiv \Delta$ is not true and δΔ is undefined. Our analysis, then, explains why we cannot always infer 'It is impossible for Epimenides the Cretan to say that nothing said by a Cretan is the case'. For, when uttered by a Cretan, the component words 'Nothing said by a Cretan is the case' themselves say nothing.

This conclusion is attractive, and it further illustrates the utility of partial logic in analysing paradoxes of this kind. It also shows, though, that there are few further formal principles to be found about when sentences say something. A sentence A will say something if and only if $A \equiv A$, but contingent liars such as the present example show that there is no *formal* test for when this condition is met. Whether $A \equiv A$ may depend on whether the speaker is a Cretan, or upon a host of other extra-linguistic factors. Here or hereabouts, formal methods give out.

6.7 Conclusion

Recalling his youthful flirtation with the so-called 'performative' analysis of the notion (a quite different account from that which we have been exploring), Strawson remembered 'fatuously announcing to [his colleague] George Paul that I had a new theory of truth; to which he sensibly and characteristically replied: "Come on now, which of the old ones is it?"' (Strawson 1998: 8). There has been no pretence of novelty here. All the same, it can be worth blowing the dust off one of the old theories, tracing its precise relationship to its rivals, and reassessing it in the light of more recent philosophy. That is what I have tried to do in this paper.[15]

References

Blamey, S. R. (2002). 'Partial Logic', in D. M. Gabbay and F. Guenther (eds.), *Handbook of Philosophical Logic*, 2nd edition, Volume 5, 261–353. Dordrecht: Kluwer.

Dummett, M. A. E. (1981). *Frege: Philosophy of Language*, 2nd edition. London: Duckworth.

Evans, M. G. J. (1977). 'Pronouns, Quantifiers, and Relative Clauses (I)', *The Canadian Journal of Philosophy* 7: 476–536.

Frege, Gottlob (1879). *Begriffsschrift: Eine der Arithmetischen Nachgebildete Formelsprache des Reinen Denkens*. Halle: Nebert.

[15] This is the text of the 2017 George Myro Memorial Lecture at the University of California at Berkeley. While illness prevented me from delivering the lecture, I am very grateful to Berkeley for its invitation, which provided a much needed deadline for getting my thoughts about truth in order. I also thank those who made comments at the places where I have presented these ideas: the Moral Sciences Club at Cambridge, the Universities of Augsburg, Glasgow, and Leeds, New York University, and my seminar at Oxford.

Halbach, V. and L. F. Horsten (2006). 'Axiomatizing Kripke's Theory of Truth', *The Journal of Symbolic Logic* 71: 677–712.

Horn, L. R. (1989). *A Natural History of Negation*. Chicago: University of Chicago Press.

Horsten, L. F. (2011). *The Tarskian Turn: Deflationism and Axiomatic Truth*. Cambridge, MA: MIT Press.

Karttunen, L. (1973). 'Presuppositions of Compound Sentences', *Linguistic Inquiry* 4: 167–93.

Kripke, S. A. (1975). 'Outline of a Theory of Truth', *The Journal of Philosophy* 72: 690–716.

Parsons, C. D. (1974). 'The Liar Paradox', *Journal of Philosophical Logic* 3: 381–412.

Parsons, T. (1984). 'Assertion, Denial, and the Liar Paradox', *Journal of Philosophical Logic* 13: 137–52.

Prior, A. N. (1958). 'Epimenides the Cretan', *The Journal of Symbolic Logic* 23: 261–6.

Prior, A. N. (1961). 'On a Family of Paradoxes', *Notre Dame Journal of Formal Logic* 2: 16–32.

Prior, A. N. (1971). *Objects of Thought*, ed. P. T. Geach and A. J. P. Kenny. Oxford: Clarendon Press.

Putnam, H. W. (1973). 'Meaning and Reference', *The Journal of Philosophy* 70: 699–711.

Ramsey, F. P. (1927). 'Facts and Propositions', *Proceedings of the Aristotelian Society Supplementary Volume* 7: 153–70. Page references are to the reprint in Ramsey, ed. D. H. Mellor, *Philosophical Papers* (Cambridge: Cambridge University Press, 1990), 34–51.

Ramsey, F. P. (1991). *On Truth*. Dordrecht: Kluwer.

Read, S. L. (2009). 'Plural Signification and the Liar Paradox', *Philosophical Studies* 145: 363–75.

Rothschild, D. (2008). 'Presupposition Projection and Logical Equivalence', *Philosophical Perspectives* 22: 473–97.

Rumfitt, I. (2011). 'Ramsey on Truth and Meaning', in B. C. A. Morison and K. Ierodiakonou (eds.), *Episteme, etc.: Essays in Honour of Jonathan Barnes*, 213–45. Oxford: Clarendon Press.

Rumfitt, I. (2015). *The Boundary Stones of Thought: An Essay in the Philosophy of Logic*. Oxford: Clarendon Press.

Rumfitt, I. (2016). 'Objects of Thought', in Gary Ostertag (ed.), *Meanings and Other Things: Themes from the Work of Stephen Schiffer*, 73–94. Oxford: Oxford University Press.

Rumfitt, I. (2017). 'The Liar without Truth', in Bradley Armour-Garb (ed.), *Reflections on the Liar*, 191–211. New York: Oxford University Press.

Russell, B. (1905). 'On Denoting', *Mind* 14: 479–93.
Scott, D. S. (1975). 'Combinators and Classes', in C. Böhm (ed.), *λ-Calculus and Computer Science*, 1–26. Heidelberg: Springer.
Stalnaker, R. C. (1974). 'Pragmatic Presuppositions', in M. K. Munitz and P. Unger (eds.), *Semantics and Philosophy* (New York: New York University Press), 197–213.
Strawson, P. F. (1950). 'On Referring', *Mind* 59: 320–44. Page references are to the reprint in Strawson 1971: 1–27.
Strawson, P. F. (1964). 'Identifying Reference and Truth-Value', *Theoria* 30: 96–118. Page references are to the reprint in Strawson 1971: 75–95.
Strawson, P. F. (1970). 'Meaning and Truth. An Inaugural Lecture delivered before the University of Oxford on 5 November 1969'. Oxford: Clarendon Press. Page references are to the reprint in Strawson 1971: 170–89.
Strawson, P. F. (1971). *Logico-Linguistic Papers*. London: Methuen.
Strawson, P. F. (1998). 'Intellectual Autobiography', in L. E. Hahn (ed.), *The Philosophy of P. F. Strawson: The Library of Living Philosophers Volume XXVI*, 3–21. Chicago: Open Court.
Tappenden, J. (1999). 'Negation, Denial and Language Change in Philosophical Logic', in D. M. Gabbay and H. Wansing (eds.), *What is Negation?*, 261–98. Dordrecht: Kluwer.
Tarski, A. (1935). 'Der Wahrheitsbegriff in den formalisierten Sprachen', *Studia Philosophica* I: 261–405. Translated by J. H. Woodger as 'The Concept of Truth in Formalized Languages' in Tarski, ed. Woodger and J. Corcoran, *Logic, Semantics, Metamathematics*, 152–278. Indianapolis: Hackett, 1983.
Tarski, A. (1944). 'The Semantic Conception of Truth and the Foundations of Semantics', *Philosophy and Phenomenological Research* 4: 341–76.
Wiggins, D. R. P. (1993). 'Putnam's Doctrine of Natural Kind Terms and Frege's Doctrines of Sense, Reference, and Extension: Can They Cohere?', in A. W. Moore (ed.), *Meaning and Reference*, 192–207. Oxford: Oxford University Press.
Williamson, T. (1998). 'Indefinite Extensibility', *Grazer Philosophische Studien* 55: 1–24.