Existence Presuppositions and Background Knowledge

PETER LASERSOHN
University of Rochester

Abstract

When a definite noun phrase fails to refer, the statement containing it is often felt to lack a
truth value, as in *The king of France is bald*. In other examples, however, the statement seems
intuitively false, and not truth-valueless: consider the case of a speaker who points at an
obviously empty chair and says *The king of France is sitting in that chair*. The difference appears to
depend on the pragmatics of verification; we know the sentence is false because the chair is
empty—the question of the existence of the king of France need not even come up. A semantics
is sketched for assigning truth values to sentences relative to information states. A sentence
containing a definite NP may be evaluated as false relative to a given information state rather
than simply truth-valueless if, after removing the information that the NP fails to refer, the
resulting information state still cannot be consistently extended to one making the sentence
true. On this assumption, existing proposals for the semantics of negation in information-state
semantics turn out to correspond to internal and external negation, respectively.

It has long been a popular idea that statements containing definite noun phrases
which fail to refer are neither true nor false. And, indeed, intuitions regarding
the truth value of sentences like the time-worn example in (1) are not at all
clear.

(1) The king of France is bald.

We find, however, that in other sorts of examples, intuitions are much sharper.
Imagine a speaker who points at an obviously empty chair and utters (2):

(2) The king of France is sitting in that chair.

Here, we find that the statement is simply false; there is little temptation to say
that it lacks a truth value.

A similar example is (3), uttered in a situation where no noise has come from
the direction of the door:

(3) The king of France is knocking on the door.

Likewise, example (4), uttered in a situation where an obviously untouched
sandwich is on the table, seems straightforwardly false:

(4) The king of France ate that sandwich.
Conversely, in the same situation, sentences (5)-(7) are straightforwardly true:

(5) The king of France is not sitting in that chair.
(6) The king of France is not knocking on the door.
(7) The king of France has not eaten that sandwich.

What makes the difference here? These examples may be reminiscent of the following examples, discussed by Strawson (1964):

(8) Jones spent the morning at the local swimming pool.
(9) The Exhibition was visited yesterday by the king of France.

Sentence (8) may express a false statement even if there is no local swimming pool, and sentence (9) may express a false statement even if there is no king of France.

Strawson suggests that truth-valuelessness may depend on articulation of a statement into topic and comment. A topic phrase may, if it fails to refer, result in a truth value gap for the statement in which it appears; but a definite noun phrase which forms part of the comment may fail to refer without causing a truth value gap. Similar ideas have been suggested more recently by Jay Atlas (1988) and others.

This idea has some attractiveness for examples like (8) and (9), and may very well be correct. It seems less plausible for examples like (2) through (7), however, so I think something more needs to be said. One need only look at the chair or the sandwich, or listen at the door, to determine that (2)-(4) are false and that (5)-(7) are true; it makes little difference whether we regard these statements as 'about' the king of France on the one hand, or the chair, door and sandwich on the other.

A somewhat different hypothesis might trace the distinction between example (1) and examples (2)-(7) to the fact that is bald is an ‘individual’ level predicate, in the terminology of Carlson (1977) and subsequent work, while is sitting in that chair is a ‘stage’ level predicate. However, this hypothesis is quickly disconfirmed by examples like (10).

(10) The king of France is on the University of Rochester faculty.

I take this sentence to be clearly false, but is on the University of Rochester faculty is an individual level predicate, not a stage level predicate.

A more likely hypothesis attributes the difference not to the semantics of the predicate, but to the pragmatics of verification. Sentence (2) seems false and sentence (5) seems true because in the situation described we can see that there is no one in the chair. This is enough to assign truth values to these sentences without the question of the existence of a king of France even coming up. Likewise, we know enough about the Rochester faculty to feel intuitively that (10) is false,
without ever having to consider the question of whether there is a king of France. In contrast, example (1) can only be judged true or false on the basis of information about the king of France himself; it cannot be verified or falsified without addressing the issue of whether there is a king of France.

There is a likelihood of misunderstanding here, so let me state very plainly and emphatically that I am not claiming that the sentence in general must be objectively verifiable (or falsifiable) in order to have a truth value. I do believe that an affirmative statement which might otherwise be judged of indeterminate truth value (because it contains a term which fails to refer) can instead be judged false, provided the context makes it possible to determine that the statement could not possibly be true regardless of whether the term has reference or not. Conversely, negative statements can be judged true in analogous circumstances. This is the difference between (1) on the one hand and (2)-(7) and (10) on the other, and has the result that (1) may be judged of indeterminate truth value if there is no king of France, even while (2)-(7) and (10) are not.

It will be useful, in formulating an analysis of these facts, to adopt a semantic framework in which truth is relativized to ‘data sets’ or ‘information states’. This approach is adopted, for example, in the framework of Data Semantics, developed in Veltman (1981) and Landman (1986), which I will adopt in broad outline here; however, it will matter little for current purposes whether we use precisely the theory advocated by Landman and Veltman, or some other framework which adopts a similar notion of truth-relative-to-a-data-set. We write ‘\(D \models \phi\)’ for \(\phi\) is true on the basis of data set \(D\), and ‘\(D \nvdash \phi\)’ for \(\phi\) is false on the basis of data set \(D\).

Data sets themselves will be assumed to be consistent sets of propositions, with the requirement that if \(\phi \in D\) then \(D \models \phi\). The possibility is left open that \(D \models \phi\) even if \(\phi \notin D\), corresponding to the intuition that a proposition may be concluded on the basis of a set of data, without being encoded directly in that set of data. We assume that \([\phi \mid D \models \phi]\) corresponds to the closure of \(D\) under some suitable consequence relation.

The possibility exists that for a given \(D\) and \(\phi\), neither \(D \models \phi\) nor \(D \nvdash \phi\). But of course this should not be taken as meaning that \(\phi\) lacks a truth value in any absolute sense; data sets may be limited in the information they encode, and \(\phi\) may just be undecidable on the basis of \(D\). Ordinarily—though perhaps not in cases of presupposition failure—\(D\) may be extended to some more complex data set \(D’\) on the basis of which \(\phi\) does evaluate as true or false.

In giving an analysis of the sentences discussed above, we will have to make use of data set revisions—that is, transitions from one information state to another—including revisions which non-monotonically ‘remove’ information. Given a data set \(D\), and assuming \(\phi\) not to be tautological, let \(D - \phi\) be the data set as much like \(D\) as possible, compatible with the condition that
We understand $D - \phi$ to be the data set which results from the removal of any commitment to $\phi$ from $D$.

Besides the removal of information from a data set, we must also consider the addition of information. In particular we will have occasion to quantify over the possible ways of extending a given data set. We write $D < D'$ for '$D'$ is an extension of $D'$, which we understand to mean simply that $D'$ is a consistent superset of $D$.

We can now return to our original examples. Why is it that someone who points at an empty chair and says *The king of France is sitting in that chair* seems to be saying something false? I would like to suggest that it is because *even if we suspend our knowledge that there is no king of France, there is no way of consistently extending our information to include the proposition that the king of France is sitting in the chair*. Such an extension is impossible because we know the chair to be empty.

In contrast, if we suspend our knowledge that there is no king of France, our information may then be extended either to include the proposition that the king of France is bald, or to include the proposition that the king of France is not bald.

This suggests a falsehood clause for definite descriptions like that in (11b), which we pair with the more ordinary Russellian truth clause in (11a):

**(11)**

a. $D \vdash (\text{the } x : \phi) \psi$ iff $D \vdash \exists y [\forall x [\phi \leftrightarrow x = y] \land \psi]$.

b. $D \vdash (\text{the } x : \phi) \psi$ iff for all $D'$ such that $(D - \exists y [\phi \leftrightarrow x = y]) < D'$, it holds that $D' \vdash \exists y [\forall x [\phi \leftrightarrow x = y] \land \psi]$.

In English: *The $P$ is $Q$* is true on the basis of data set $D$ just in case it is true on the basis of $D$ that there is a unique $P$, which is $Q$. *The $P$ is $Q$* is false on the basis of $D$ iff after removing from $D$ the information that there is no unique $P$, the resulting data set still only extends to ones on the basis of which it is false that there is a unique $P$, which is $Q$.

Taking this approach has an interesting consequence. Supposing that we can never know that the king of France would not be bald if there were one (perhaps a questionable supposition), *The king of France is bald* will come out truth-valueless relative to any data set encoding the information that there is no king of France—even in data sets which are 'total' in the sense of not allowing any proper extensions. Thus we have a truth value gap that represents something more than simple incompleteness of information.

Aside from its persistence into total information states, however, this truth value gap has the same status as more run-of-the-mill gaps of the sort that have nothing to do with presupposition, but only with the general relativization of truth to data sets. In the context of this sort of semantic framework, it is perhaps not best to think of an individual sentence as 'coming with' presuppositions,
which must be met if the sentence is not to be sapped of the truth value it would otherwise have; rather, the gap represents a kind of residue of undecidability, left even after all possible ways of extending a data set are considered.

In this context, it is worth considering the interaction of definites and negation. Veltman (1981) suggests the following clauses for negation in Data Semantics:

(12) a. $D \vdash \neg \phi$ iff $D \vdash \phi$

b. $D \vdash \neg \phi$ iff $D \vdash \phi$

Combined with the rules given above for definites, these clauses appear to give the desired results: A sentence like The king of France is not sitting in that chair will be assigned the value true in a context where it is known that the chair is empty; but The king of France is not bald will not be assigned a truth value—assuming it is known that there is no king of France, and in the absence of information which would settle the baldness issue if this knowledge were suspended.

It is interesting to contrast the clauses in (12) with the one given in (13), essentially that used in Kripke-style semantics for Intuitionistic Logic:

(13) $D \vdash \neg \phi$ iff for all $D' \supseteq D$, $D' \not\models \phi$.

Here, as with (12), The king of France is not sitting in that chair is assigned the value true in the relevant context. However, The king of France is not bald comes out not truth-valueless, but true. I take this not to be the correct result in the general case. However, it is possible to view (13) as giving the semantics for so-called 'external', presupposition-cancelling negation, and (12) as giving the semantics of the more usual 'internal' presupposition-preserving negation. This idea is attractive, but I doubt that it really tells the whole story on this kind of negation. More likely, presupposition-cancelling negation is just one instance of the much more general phenomenon of 'metalinguistic' negation, as argued by Horn (1985).

It may be worth contrasting the analysis given here with that of Lappin and Reinhart (1988), which is similar in certain respects. Like the proposal presented here, Lappin and Reinhart's analysis attempts to account for speaker intuition of a truth value gap by appealing to the pragmatics of verification. Unlike the present proposal, Lappin and Reinhart's analysis makes heavy use of generalized quantifier theory, and is stated without explicit appeal to partial information states.

On Lappin and Reinhart's view, the apparent truth value gap which results when a definite noun phrase fails to refer is just one instance of a more general phenomenon which occurs with all strong determiners. For example, speakers will regard sentence (14) as of indeterminate truth value, given that there have been no American kings:
Existence Presuppositions and Background Knowledge

(14) All American kings lived in New York.

In contrast, analogous sentences with weak determiners, e.g. (15), will normally be judged false in similar circumstances:

(15) Five American kings lived in New York.

Lappin and Reinhart suggest that sentences containing weak determiners differ from sentences containing strong determiners in how they may be assessed for truth or falsity, and it is this difference in assessment procedures which results in the intuition of a truth value gap for examples like (14).

The truth or falsity of a sentence containing a weak determiner in its subject noun phrase may be assessed simply by checking the cardinality of the intersection of extension of the determiner’s N’ argument with the extension of its VP argument. For example, the truth or falsity of example (15) may be assessed simply by checking the cardinality of the intersection of the set of American kings with the set of things that lived in New York. If this intersection is of cardinality 5 or greater, the sentence is true; if it is of cardinality less than 5, it is false. Note that the cardinality of the extension of the N’, the set of all American kings, need not be checked.

In contrast, according to Lappin and Reinhart, in order to assess the truth or falsity of a sentence with strong determiner, one must check the extension of the determiner’s N’ argument. Sentence (14), for example, cannot be assessed for truth or falsity simply by checking the set of American kings that lived in New York; one must check the set of all American kings.

According to Lappin and Reinhart, the fact that one must check this empty set is what results in the intuition of a truth value gap. In their words: ‘whenever the assessment of a sentence must start with a scan of an N’ set of a given NP, assessment is stalled if this set is empty. In this case, the sentence is marked as anomalous, empirically irrelevant, or undefined, regardless of its semantic interpretation.’ Since the assessment of the sentence as true or false ‘stalls’, speakers will not have clear intuitions as to whether the sentence is true or false.

Lappin and Reinhart’s proposal can be attacked from several angles. First, it is stated entirely in terms of the logical properties of determiners. For that reason it would not appear to extend naturally to determinerless examples—those involving proper names, for instance. However, proper names behave much like definite descriptions with regard to reference failure and the intuition of truth value gaps. Given that the legend of Santa Claus does not specify his exact weight, sentence (17a) does not seem intuitively true or intuitively false. However (17b), uttered by someone pointing at an obviously empty chair, is clearly false:

(17) a. Santa Claus weighs exactly 275 pounds.
   b. Santa Claus is sitting in that chair.
The basic approach suggested above, however, would seem to extend to such examples fairly straightforwardly. Example (17b) will be assigned the value false in the context given, because even if we suspend our knowledge that there is no Santa Claus, we cannot consistently extend our data set to one where he is sitting in the chair—we can see that the chair is empty. But once we suspend our knowledge that there is no Santa Claus, it seems that we could equally well extend our data set to one where he weighs exactly 275 pounds, or one where he doesn't.

Lappin and Reinhart's proposal can also be attacked on more technical grounds. As they themselves point out (following an anonymous reviewer), it is not really the case that the truth value of a sentence containing a universal determiner can only be assessed by checking the extension of the determiner's N' argument. Instead, a sentence of the form 'Every A is a B', for example, can be assessed by scanning the set A-B. Given the standard generalized quantifier semantics for universal determiners, if this set is empty, the sentence is true; one need not check the entire set A at all. Since one need not check the entire N' set of the determiner, there is no point at which the assessment procedure must stall if this set is empty.

Lappin and Reinhart attempt to meet this objection by asking us to consider examples such as Every unicorn is intelligent. They claim that if one attempts to assess this sentence by checking A-B (that is, [unicorn] \{intelligent\}), 'the speaker will have to represent and scan the set of things that are not intelligent. From the perspective of computational efficiency, this is the least efficient way to assess the sentence.' Hence, this procedure will not be easily available. However, Lappin and Reinhart's analysis would seem to require precisely this procedure for examples like Five unicorns are unintelligent, so it is not at all clear that this explanation can go through.

We may also object to Lappin and Reinhart's proposal on the grounds that it simply is not clear why 'scanning' an empty set should cause the assessment procedure to stall. In fact it is clear that Lappin and Reinhart must allow the empty set to be successfully scanned in certain circumstances. For example, to assess the truth or falsity of a sentence like (18), on their view one scans the set of unicorns sitting in the chair. If it is empty, the sentence is judged to be true; no stall of the assessment procedure takes place.

(18) No unicorn is sitting in that chair.

In fact, it may be that nothing is sitting in the chair, in which case the extension of the N', the extension of the VP, and their intersection will all be empty. Yet the sentence is trivially easy to assess for truth. Apparently, it is only the N' set of a determiner which causes a stall if it is empty; otherwise the empty set may be scanned without problems. This is a rather surprising asymmetry, and it is mysterious why it should obtain.
Finally, we may note that Lappin and Reinhart's analysis does not allow the kind of sensitivity to context and background knowledge that the analysis suggested here does. Their analysis does not predict that sentences like *The king of France is sitting in that chair* will be judged as false in the context given, for example. Since there is no king of France, the N' set of the determiner *the* is empty. In Lappin and Reinhart's view, this set must be scanned to assess the sentence for truth or falsity, since *the* is a strong determiner. But if the N' set is empty, then scanning it should result in an assessment stall, and the sentence should be judged of undefined truth value. But this result is incorrect; the sentence seems clearly false. I think we can conclude that in certain circumstances, the truth or falsity of a sentence containing a strong determiner can be assessed without 'scanning' the determiner's N' set, and that an analysis which crucially assumes that this set must be scanned is inadequate.

To summarize, we find that certain sentences containing non-referring definite descriptions are clearly false, while others do not provoke clear intuitions of either truth or falsity. The crucial difference seems to be that, in cases of clear falsity, the context provides an independent reason to believe that the sentence cannot be true, regardless of the question of whether the definite succeeds in referring or not. To account for this, I sketched a semantics in which truth values are assigned relative to data sets. The truth value (relative to a given data set) of a sentence containing a definite description was analysed as depending in part on what information was supported even if the knowledge that the definite fails to refer is suspended. This yields a system in which reference failure sometimes results in truth-valuelessness and sometimes results in falsehood (for affirmative sentences). This proposal was contrasted with one suggested by Lappin and Reinhart (1988), which was shown to be problematic in several respects.

**Acknowledgements**

Thanks to three anonymous referees, and to Louise McNally and Peter Svenonius, who practically deserve co-author status, but should not be held responsible for errors.

PETER LASERSONH  
*Department of Foreign Languages, Literatures  
and Linguistics  
University of Rochester  
390 Dewey Hall  
Rochester, NY 14627  
USA*
NOTES

1 Nor am I claiming that sentences must be verifiable in order to be meaningful, or that the rules which assign truth conditions to sentences must be stated in terms of the procedures by which we verify that a sentence's truth conditions are satisfied, or anything of the sort.

2 Relativization of truth to information states is not unique to Data Semantics, of course, and goes back at least to Kripke's (1965) semantics for Intuitionistic Logic.

3 This is not to claim that we must take the consequence relation as antecedently given and define \( \phi \mid D \rightarrow \phi \) in terms of it. Rather, one can give a recursive definition of the notions of truth and falsity on the basis of \( D \), and then define the consequence relation in terms of these data-oriented notions of truth and falsity.

4 A rigorous definition of \( D - \frac{\phi}{\phi} \) is not a trivial enterprise. See, e.g., Gardenfors (1988) for relevant discussion.

5 Strong determiners are those which cannot appear in postcopular position in an existential there construction, and include the, every, all, most, neither, and related determiners:

\[
\begin{align*}
\text{(i) } & \text{a. "There is the unicorn in the garden.}\nonumber \\
& \text{b. "There is every unicorn in the garden.}\nonumber \\
& \text{c. "There are all unicorns in the garden.}\nonumber \\
& \text{d. "There are most unicorns in the garden.}\nonumber \\
& \text{e. "There is neither unicorn in the garden.}\nonumber \\
\text{(ii) } & \text{a. There is a unicorn in the garden.}\nonumber \\
& \text{b. There are some unicorns in the garden.}\nonumber \\
& \text{c. There are many unicorns in the garden.}\nonumber \\
& \text{d. There are no unicorns in the garden.}\nonumber \\
& \text{e. There are five unicorns in the garden.}\nonumber 
\end{align*}
\]

For relevant discussion see Milsark (1977), Barwise and Cooper (1981).

REFERENCES


Carlson, Greg (1977), 'Reference to kinds in English', University of Massachusetts dissertation.


