Lecture 36
Lewis's Definition of Theoretical Terms

Patrick Maher

Scientific Thought II
Spring 2010

- The Ramsey sentence of T says T is realized:

$$\exists x_1 \ldots x_n \top [x_1 \ldots x_n]$$

- The Carnap sentence says if T is realized then T is true:

$$\exists x_1 \ldots x_n \top [x_1 \ldots x_n] \supset \top [\tau_1 \ldots \tau_n]$$

- According to Carnap:
  - The Ramsey sentence contains the factual part of T.
  - The Carnap sentence specifies the interpretation of the T-terms.

It is important to separate three cases. T may have precisely one realization, or no realization, or more than one realization. (431)

T has precisely one realization (431)

In case T is uniquely realized, the Carnap sentence clearly gives exactly the right specification. It says that the T-terms name the entities in the n-tuple that is the unique realization of T. The first T-term, τ_1, names the first component of the unique realization of T; τ_2 names the second component; and so on.

T has no realization (432)

In case T is not realized, the Carnap sentence says nothing about the denotation of the T-terms. But this modesty seems to be uncalled for. The T-terms were introduced on the assumption that T was realized, in order to name the components of a realization of T. There is no realization of T. Therefore they should not name anything. "Phlogiston" presumably is a theoretical term of an unrealized theory; we say without hesitation that there is no such thing as phlogiston. What else could we possibly say? Should we say that phlogiston is something or other, but (unless phlogiston theory turns out to be true after all) we have no hope of finding out what?

T has more than one realization (432)

There remains the case in which T is multiply realized. In this case, the Carnap sentence tells us that the T-terms name the components of some realization or other. But it does not tell us which; and there seems no nonarbitrary way to choose one of the realizations ... [This doesn't do] justice to our naive impression that we understand the theoretical terms of a true theory, and without making any arbitrary choice among realizations. We should not accept Carnap's treatment in this case if we can help it.

Lewis goes on to argue that we should say multiply realized theories are false and their T-terms don't refer to anything.

Definition (437–38, simplified)

- If T has a unique realization, then τ_i denotes the ith entity in that realization.
- If T doesn't have a unique realization, then τ_i doesn't denote anything.

The sense/reference distinction (435)

- The *reference* or *denotation* of a term is what the term names in the actual world.
- The *sense* or *meaning* of a term determines what the term names in every possible world.

Examples (not in Lewis)

"Species with a heart" and "species with a kidney":

- Have the same reference/denotation.
- Have different senses/meanings.

"Bachelor" and "unmarried man":

- Have the same reference/denotation.
- Have the same sense/meaning.

The definition of T-terms gives their sense (435)

- Different possible worlds differ in whether T has a unique realization.
- According to Lewis's definition:
  - In worlds in which T does have a unique realization, τ_i denotes the ith element in that realization.
  - In worlds in which T doesn't have a unique realization, τ_i doesn't denote anything.
- So Lewis's definition says what τ_i denotes in every possible world.
- Hence Lewis's definition specifies the sense of T-terms.
- To determine the reference of T-terms, we need to know which world we are in; that requires empirical investigation.

So far, we have discussed the interpretation of T-terms only at the time of their introduction, the time when the parent theory T is first proposed. It remains to ask what happens later when T is amended and extended. (445)

If use the current version of T (445)

We might say that the T-terms should always be defined using the currently accepted version of T. As T is corrected, modified, extended, or perhaps even when we accept miscellaneous hypotheses that contain T-terms but do not belong integrally to any version of T, the T-terms gradually change their meaning . . . But these are very peculiar changes of meaning—so peculiar that this position seems to change the meaning of "change the meaning of." They occur continually, unnoticed, without impeding communication.

If use the original version of T (446)

- *We might therefore prefer to say that the T-terms keep the meanings they received at their first introduction. They should still be defined using the original version of T even after it has been superseded by revised versions.*

- *This position will work only if we permit T-terms to name components of the nearest near-realization of T, even if it is not a realization of T itself. For after T has been corrected, no matter how slightly, we will believe that the original version of T is unrealized.*

- *According to this position, we may be unable to discover the meanings of theoretical terms at a given time just by looking into the minds of the most competent language-users at that time. We will need to look at the past episodes of theory-proposing in which those terms were first introduced into the language ... If we were ignorant of history, we could all be ignorant or mistaken about the meanings of words in common use among us.*

Neither solution is attractive (446)

I do not wish to decide between these alternatives. Either seems defensible at some cost. I hope the truth lies in between, but I do not know what an intermediate position would look like.

1. Does Lewis agree that the Carnap sentence of a theory T gives the meaning of its T-terms? If not, why not?

2. State Lewis's definition of the T-terms of a theory. Does this definition specify the meaning of the T-terms, or does it merely say what they designate? Justify your answer to the latter question.

3. When Lewis considers what happens when a theory T is modified, he considers two possibilities:
   (a) The T-terms are defined using the revised version of T.
   (b) The T-terms are defined using the original version of T.

   Explain why neither solution is attractive.

Reference

📄 David Lewis.
How to define theoretical terms.
Journal of Philosophy, 67:427–446, 1970.
Online at JSTOR free with a uiuc connection.
Reprinted in David Lewis, *Philosophical Papers*, vol. 1.
Numbers in parentheses are page numbers of the original article.

📄 David Papineau.
Theory-dependent terms.
Philosophy of Science, 63:1–20, 1996.
Online at JSTOR free with a uiuc connection.
Proposes a solution to the dilemma at the end of this lecture.