

Philosophy 5340 – Epistemology

Topic 3: Analysis, Analytically Basic Concepts, Direct Acquaintance, and Theoretical Terms

Part 3: David Lewis's "How to Define Theoretical Terms"

Introductory Section

1. Here Lewis starts by advancing the following claims:
 - (1) There is a general method for defining theoretical terms.
 - (2) Reduction of one theory to another does not involve **independent** bridge laws.
2. As Lewis points out, what he is doing is not best described as **eliminating** theoretical terms. It is best described as a **vindication** of theoretical terms: "for to define them is to show that there is no good reason to want to do without them." (427)
3. Lewis's goal is not to define theoretical terms within an observation language.
4. Lewis contends that the idea of there being a part of the language that is observational, and non-theoretical, is obscure.

Comments

- (1) This is to hold, it seems to me, that there are **no analytically basic terms or concepts**.
- (2) The latter, in turn, is only plausible, I suggest, if one thinks – as Lewis does – that there are no qualitative properties, no qualia, of which one is directly aware.

5. Consequently, rather than phrasing things in terms of a distinction between observational terms and theoretical terms, Lewis suggests that one need merely distinguish between terms that are already understood – **O-terms** – and terms that are introduced by the theory under consideration – **T-terms**.
6. Lewis's goal is not to dispense with theoretical entities:

"I am also not planning to 'dispense with theoretical entities'. Quite the opposite. The defining of theoretical terms serves the cause of scientific realism. A term correctly defined by means of other terms that admittedly have sense and denotation can scarcely be regarded as a mere bead on a formal abacus. If it purports to name something, then if the theory that introduced it is true it does name something." (428)
7. Theoretical terms can name **observable entities**. Consider " H_2O ".
8. **Theoretical entities** can be named by expressions constructed from observational vocabulary. Consider "living creature too small to see".

The Postulate of *T*

1. **The postulate of *T*** can be taken to be a single sentence, since it can be taken as the conjunction of all of the sentences that make up theory *T*.

- 2. The T -terms of a theory T are the new terms that T introduces. The old terms that theory T contains are the O -terms.
- 3. The T -terms are all to be taken as **names**, rather than as **predicates** or **functors**.

Comments

(1) Typically, the T -terms of a theory will contain **predicates**. Suppose ' P ' is such a predicate. Then Lewis's idea is that we can replace sentences of the form ' Px ' by sentences of the form ' x has the property of P -hood'. In a sentence of the form ' Px ', ' P ' is a predicate, whereas in sentences of the form ' x has the property of P -hood', the expression ' P -hood' is a name of a property.

(2) Similarly, the T -terms of a theory will often contain **functors**. This will normally be the case if the theory in question is a **quantitative** theory. So consider, for example, a simple theory that consists of Newton's Laws of Motion and the Newtonian Law of Gravitation. That theory will contain a term – let us use the term ' M ' – that is connected with the idea of mass, and usually it will enter into sentences of the form ' $M(x) = y$ ', where x is some object, and y is a number that specifies the mass of object x .

Lewis's idea in this case is that we can replace sentences of the form ' $M(x) = y$ ' by sentences of the form ' x is mapped into y by the M -function'. In a sentence of the form ' $M(x) = y$ ', ' M ' is a functor, whereas in sentences of the form ' x is mapped into y by the M -function', the expression ' M -function' is a name of a function.

(3) Why is this important? The answer is that ordinary quantification involves replacing the **name** of something by a variable, as in the following inference:

$$\underline{Fa}$$

Therefore: $(\exists x)Fx$

Similarly, if one has a sentence of the form ' a has property P ', one can move, via existential generalization to ' $(\exists x)(a \text{ has property } x)$ '.

(4) By contrast, if one instead used sentences of the form ' Pa ', and moved from that to a sentence of the form $(\exists F)Fa$, one would be using what is called second-order quantification, and this raises unnecessary problems.

(5) A metaphysical footnote: What is a **function**, as the term is being used in this context? Evidently, as the example of mass illustrates, functions are related to **quantitative properties**. But what are quantitative properties? The functional representation suggests that they are mappings from particulars into numbers. Such mappings are, of course, just certain sorts of **relations**. However viewing quantitative properties in this way commits one to the conclusion that scientific theories of the relevant sorts are ontologically committed to the existence of **numbers**. If so, one can prove that numbers exist by confirming certain scientific theories.

Many philosophers have felt that one should avoid this by reformulating any such scientific theories so that there is no quantification over numbers. Perhaps the most notable defense of this view is that set out by Hartry Field in his book *Science without Numbers*.

4. Lewis assumes that all the occurrences of the T -terms in T are in **extensional contexts**, so that **existential generalization is permissible**, and also substitution on the basis of identities.

Comments

(1) In a non-extensional context, existential generalization is not a valid operation. So, for example, the following inference is invalid:

John worships Zeus

Therefore: $(\exists x)(\text{John worships } x)$

(2) Existential quantification, however, lies at the heart of the method of defining theoretical terms that Lewis is setting out, so one must insist that all T -terms occur only in extensional contexts.

(3) But what would one do if one were dealing with, say, a theory concerning mental states – such as beliefs – where one did have theoretical terms occurring within **non-extensional contexts**? The answer is that one would need to have entities such as **concepts** or **propositions** in one's ontology, in order to translate the sentences containing non-extensional contexts into sentences containing only extensional contexts.

5. Finally, one needs to assume that theory T is false if **any** of the T -terms fails to refer to anything. (The fact that a sentence contains a term that fails to refer to anything does not in general make a sentence false. Consider, for example, the sentence "There are no unicorns." But the T -terms of a theory are supposed to be referring to properties, relations, functions, etc.)

6. Given that all of the T -terms of theory T must refer to something if theory T is true, if T_1 is any T -term, then theory T will entail:

$(\exists x)(x = T_1)$.

7. The problem at this point is that in standard systems of logic, an identity statement such as ' $a = a$ ' is generally viewed as a **logical truth**, but it also entails, by existential generalization, $(\exists x)(x = a)$. But any statement entailed by a logical truth must itself be a logical truth, and so $(\exists x)(x = a)$ must be a logical truth. But we want sentences of the form ' $(\exists x)(x = T_1)$ ', where ' T_1 ' is a T -term, to express a contingent truth.

8. One possible solution is to adopt what is called a 'free logic' – a type of system of logic developed by Dana Scott – in which terms that fail to refer to anything are admissible. In such a system, ' $a = a$ ' will be true even if ' a ' fails to refer to anything, but the inference from sentences of the form ' $a = a$ ' to sentences of the form ' $(\exists x)(x = a)$ ' is not a valid inference.

9. I think that there are objections to such an approach, but this is an issue we cannot pursue here.

The Ramsey and Carnap Sentences of T

1. Lewis rewrites theory T in such a way that it is explicit what T -terms it contains:

$T[T_1, T_2, \dots T_n]$.

2. Next, there is what he refers to as the **realization formula** of T , which is obtained by replacing all of the T -terms in a uniform way by variables that do not already occur in theory T : $T[x_1, x_2, \dots, x_n]$.

3. Any n -tuple of entities – that is, any ordered set containing n entities – that satisfies the realization formula will be described as a **realization** of theory T .

4. Next, Lewis introduces the **Ramsey sentence** of theory T . This is the sentence that results when existential quantifiers involving the n variables x_1, x_2, \dots, x_n are placed in front of the realization formula:

$$(\exists x_1)(\exists x_2) \dots (\exists x_n)T[x_1, x_2, \dots, x_n].$$

The Ramsey sentence asserts, then, that **there is some** realization of theory T .

5. Finally, there is the **Carnap sentence** of theory T :

$$(\exists x_1)(\exists x_2) \dots (\exists x_n)T[x_1, x_2, \dots, x_n] \supset T[T_1, T_2, \dots, T_n].$$

The Carnap sentence says that if there is at least one realization of theory T , then the n -tuple of entities named by the n T -terms T_1, T_2, \dots, T_n is a realization of theory T .

6. It is possible to prove the following results:

(1) Every **O-sentence** that is entailed by the theory T is also entailed by the Ramsey sentence $(\exists x_1)(\exists x_2) \dots (\exists x_n)T[x_1, x_2, \dots, x_n]$ (and conversely).

(2) Any O -sentence that is entailed by the Carnap sentence $(\exists x_1)(\exists x_2) \dots (\exists x_n)T[x_1, x_2, \dots, x_n] \supset T[T_1, T_2, \dots, T_n]$ is a **logical truth**.

7. David Lewis sums up the situation as regards the Ramsey and Carnap sentences as follows:

“Therefore, insofar as theory T serves as a device for systematizing O -sentences, the Ramsey sentence of T will do the job as well as the postulate itself. That was Ramsey’s observation. The Ramsey sentence can obviously do nothing to help interpret the T -terms, since they do not occur in it. The Carnap sentence, on the other hand, does nothing to help systematize O -sentences; but it does contain the T -terms, and it does seem to do as much toward interpreting them as the postulate itself does. And the Ramsey and the Carnap sentences between them do exactly what the postulate does.”

“Accordingly, Carnap proposes to take the Ramsey sentence as the **synthetic postulate** of T and the Carnap sentence as the **analytic postulate** of T . They divide the labor of the original postulate, which both systematized O -sentences and partially interpreted the T -terms.” (Lewis, 431)

The Interpretation of T -Terms

1. Lewis will argue in this section that the Carnap sentence does **not** “specify the appropriate interpretation of the T -terms, insofar as appropriate interpretations can be specified at all.” (431)

2. Lewis points out that we need to consider three cases:

(1) The case where there is **one and only one realization** of theory T .

(2) The case where there is **no realization** of theory T .

(3) The case where there is **more than one realization** of theory T .

3. In the case where there is one and only one realization of theory T – that is, one and only one n -tuple of entities that satisfy that realization formula $T[x_1, x_2, \dots, x_n]$, the Carnap sentence $(\exists x_1)(\exists x_2) \dots (\exists x_n)[x_1, x_2, \dots, x_n] \supset T[T_1, T_2, \dots, T_n]$ correctly says that the n -tuple (T_1, T_2, \dots, T_n) is the n -tuple that satisfies the theory in question.

4. In the case where there is no realization of theory T , the Carnap sentence says nothing about the denotations of the T -terms. Lewis argues that this is unsatisfactory, and that what a sound account will do is to say that the T -terms **do not have any denotation** in such a case.

5. Finally, there is the case where there are multiple realizations of theory T . Here Lewis says,

“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 to be no nonarbitrary way to choose one of the realizations. So either the T -terms do not name anything, or they name the components of an arbitrarily chosen one of the realizations of T . Either of these alternatives concedes too much to the instrumentalist view of a theory as a mere formal abacus. Neither does justice to our naïve 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. Can we?” (432)

6. The crucial question that arises is whether the following thesis is true:

Every true theory has more than one realization – that is, there is more than one n -tuple that satisfies the realization formula.

If this thesis is true, we cannot improve on what the Carnap sentence does.

7. Lewis refers to a proof set out by John Winnie that “scientific theories cannot be uniquely realized.” Lewis’s response to Winnie’s proof is this: “I am concerned only with realizations under a fixed interpretation of the O -vocabulary; whereas Winnie permits variation in the interpretation of certain O -terms from one realization to another, provided that the variation is confined to theoretical entities. For instance, he would permit variation in the extension of the O -predicate ‘__ is bigger than __’ so long as the extension among observational entities remains fixed. Winnie’s proof does not show that a theory is multiply realized in my sense unless the postulate of the theory is free of ‘mixed’ O -terms: O -predicates whose extension includes theoretical entities, and the like. I would claim that mixed O -terms are omnipresent, and that there is no reason not to grant them a fixed interpretation even as applied to theoretical entities.” (433-4)

Comments

(1) Lewis’s response to Winnie is made against the background of Lewis’s belief that there is not really any **rock-bottom observational language**. How do things look if one rejects Lewis’s view here?

(2) If one accepts the idea of analytically basic terms that pick out properties and relations with which one can be directly acquainted, or of which one is directly aware, some of those terms – namely, those that pick out qualitative **properties** – will not be terms that are true of theoretical entities. But among the analytically basic terms, there

will also be ones that pick out **spatial relations** – such as **the relation of betweenness** – and those relations will also hold between theoretical entities.

(3) In addition, **causation** can be defined so that it picks out a unique relation, and it is a relation that holds between **theoretical states of affairs** as well as observational states of affairs.

8. Given that it seems possible for there to be theories that have exactly one realization, Lewis concludes that we are justified in holding that “*T*-terms ought to name the components of the unique realization of *T* if there is one, and ought not to name anything otherwise.” (434)

9. If this is right, we can lay down the following three postulates:

(1) If *T* is uniquely realized, then the *T*-terms T_1, T_2, \dots, T_n denote, in the relevant order, the *n*-tuple of entities that is a realization of theory *T*.

(2) If *T* is not realized, then the *T*-terms T_1, T_2, \dots, T_n do not denote anything.

(3) If *T* is multiply realized, then the *T*-terms T_1, T_2, \dots, T_n do not denote anything.

10. Lewis interprets these postulates in such a way that the denotations of *T*-terms can vary from one possible world to another. So a theoretical term such as “electron” might pick out one property in the actual world, but a different property in another possible world. He is, in other words, interpreting the first of his three postulates as follows:

(1) If *T* is uniquely realized in world *W*, then the *T*-terms T_1, T_2, \dots, T_n do denote things in world *W*, and what they denote, in the relevant order, is the members of the unique *n*-tuple of entities **in world *W*** that realize *T* in **that** world.

11. It is possible, however, to re-interpret Lewis’s first postulate in the following way:

(1*) If *T* is uniquely realized in the **actual** world, then the *T*-terms T_1, T_2, \dots, T_n denote things in world *W* if and only if the entities that the *T*-terms T_1, T_2, \dots, T_n denote in the **actual** world **also exist** in world *W*, and realize *T* in that world.

According to Lewis’s proposal, whether *T*-terms denote anything in a different possible world, and if so, what, depends only upon what exists in that other possible world, and does not depend at all on what exists in **this** world. According to the alternative proposal, if *T*-terms denote anything in another world, they denote exactly the properties and relations that they denote in **this** world.

12. A way of describing this difference is to say that on the second way of assigning meaning to theoretical terms one is either using definitions that implicitly involve **indexical terms (demonstrative terms)** referring to the actual world, or one is using the concept of the actual world. So, for example, one could say that negative charge is the third member of the unique *n*-tuple of properties, relations, and functions that realize theory *T* in **this** world, or in the actual world. Lewis’s approach, by contrast, uses sentences that are free of indexicals, and of the concept of the actual world.

13. Lewis points out that on his approach, “*T*-terms purporting to name properties will normally turn out to be logically indeterminate names of properties” (436) – that is, they

will name **different** properties in various possible worlds. Consequently, the **sense** of T -terms will often not be a **constant function**.

The Definitions of T -Terms

14. Given the three postulates that Lewis has defended, T -terms can be defined as follows:

$T_i = \text{def.}$ The i th member of the unique n -tuple of things (properties, relations, and functions) that realize theory T .

15. Lewis points out that for these definitions to be necessarily true, we need an approach that treats **denotationless** names in the way that Dana Scott's approach does:

"Under Russell's theory of descriptions, the definition sentences are disguised existential quantifications, false unless T is uniquely realized. Under the truth-value-gap theory of Frege and Strawson, atomic contexts of denotationless names or descriptions have no truth value; so the definition sentences are neither true nor false unless T is uniquely realized. Either way, the definition sentences would not be valid identities, true whether or not T is uniquely realized." (438)

The Expanded Postulate of T

16. Given the definitions of theoretical terms, one can then use those definitions to eliminate the T -terms from theory T . The result will be a sentence that is logically equivalent to a sentence that says, so to speak, that **there is a unique realization of theory T** . ("So to speak" because the sentence in question will refer not to theory T , but only to the realization formula of theory T .) Lewis refers to the definitionally expanded proposition as "**the expanded postulate of T** ".

17. The expanded postulate of T will be, in Lewis's terminology, **an O -sentence**, since it contains only the old vocabulary.

Comment

It will not, of course, be an **observational** sentence, since its truth may very well depend upon the existence of properties and relations that are in no sense observable.

18. The Ramsey sentence of theory T will also be an O -sentence, but it will be a weaker one, since the Ramsey sentence says only that **there is at least one realization of theory T** – thereby leaving it open whether there is **only one**.

19. At this point a **possible objection** to Lewis's account arises:

(1) It can be proved that any O -sentence that is entailed by T is entailed by the Ramsey sentence of T .

(2) But if the definitions of the T -terms are part of theory T , then theory T entails an **O -sentence** that is not entailed by the Ramsey sentence – **namely, the expanded postulate of T** , which entails that there is only one realization of T .

(3) So the definitions are giving us " **O -theorems that could not have been derived without them.**" (439, emphasis added)

(4) Consequently, Lewis's definitions of the theoretical terms of T are not logically implied by the postulate of theory T . The Carnap sentence, in contrast, is logically implied.

(5) The contention then is that the definition sentences of T can be correct definitions if and only if the following is true:

The theorems of T are all and only the logical consequences of the postulate of T .

(6) If his is right, then the definition sentences of T that Lewis has advanced cannot be correct.

20. Lewis's response to this objection is as follows:

"I contend that the theorist who proposed T by asserting the postulate of T explicitly, labeling it as the postulate of a term-introducing theory, has also implicitly asserted the definition sentences of T . I contend that his audience *will* take him to have implicitly asserted the definition sentences of T ." (439)

21. Lewis notes that when a number of theoretical terms T_1, T_2, \dots, T_n are introduced by a theory T , none of those terms can be true of anything unless theory T is true, since each term is defined in terms of the relevant member of an n -tuple that realizes the theory in question.

22. Lewis also notes that because of this, if T is any theory that involves laws of nature, any statement that entails that at least one of the theoretical terms in question truly applies to something will entail the existence of the relevant laws of nature, and because of this, some revision is needed in Hempel's covering-law analysis of scientific explanation. (440-1)

Derived Bridge Laws for T

23. Lewis introduces the idea of a **reduction premise** for a theory T . A reduction premise will be one that says that there are terms R_1, R_2, \dots, R_n that belong to the observational vocabulary of some other theory T^* such that the ordered n -tuple of entities picked out by the terms R_1, R_2, \dots, R_n realizes theory T .

24. Lewis points out: "A reduction premise for T does not imply the postulate of T either logically or definitionally; or conversely." (442) To connect the reduction premise for T to the postulate of T , one needs a set of n bridge laws of the form $R_i = T_i$.

25. What is the status of the bridge laws? Lewis rejects the common view that they are "separate empirical hypotheses, independent of the reducing theory T^* ." (442) Lewis suggests that this common view may reflect the fact that the bridge laws contain the theoretical terms of theory T , since this fact may suggest that the bridge laws cannot be derived from the reducing theory T^* , given that T^* does not contain the theoretical terms of theory T . But one can use the definitions of the latter terms to eliminate them from the bridge laws, and the **definitionally expanded bridge laws** will then be O^* -sentences, since we are assuming that all the O -terms of theory T are O^* -terms of the reducing theory T^* .

26. The upshot is that if theory T^* entails the relevant O^* -sentences, it also entails the bridge laws. In that case, T can be reduced to T^* **without the need of any extra, independent assumptions.**