First-year Proseminar AY '07–'08: Lecture 4

BENJ HELLIE

September 30, 2007

1 Carnap's central philosophical aims in *Meaning and necessity*

Essence and accident

Philosophical tradition, and common sense, distinguish *essence* and *accident*.

Roughly speaking, the essential features of a thing are those which constitute its nature, or define the thing, or explain what kind of thing it is. The accidental features are those which do not.

Suppose that o is F. A central test for whether Fness is accidental to o is whether o has Fness contingently: whether there are possible circumstances such that, if those circumstances held, o would exist, but lack Fness. If so, then Fness is surely not essential to o: after all, if it were, then o could have a different nature or essence than it actually has. But that doesn't really make any sense: if Fness really is definitional of o, anything that wasn't F wouldn't be o, but would be something else.

- If o has Fness contingently, Fness is accidental to o.

Should we endorse the converse? (If o has Fness necessarily, Fness is essential to o.) For a while "essential property" and "necessary property" were used equivalently, though these days many philosophers doubt this equation, for reasons that needn't detain us.

Whether things have natures "objectively" is a central issue in metaphysics: if not, then perhaps there's no real order to the world. Or any time in philosophy when we try to answer a "what is it to be an F" question, we are presupposing that Fs have a certain essence, the answer to the question. For instance, in ethics, sometimes people try to derive ethical theses from claims about the essence of action or personhood; in philosophy of mind, sometimes people say that to be conscious is to be intentional in a certain way. Taxonomic science clearly tries to say what the essences of things are; fundamental explanatory science with its exceptionless covering laws also only provides explanations against the background assumption that fundamental things can't change their categories; etc.

So a lot hangs on whether the essence/accident distinction makes sense.

Some amateur epistemology

Extreme empiricists are skeptics about a priori justification. They think that one is justified in judging that p iff one's perceptual experience has p as part of its content.

Say that p is *experience-transcendent* iff no experience could have either p or $\neg p$ as part of its content. If one is justified in accepting such a p, one is justified (in part) a priori in accepting it.

Often, even when the claim that o is F is not experience-transcendent, the claim that o is essentially F is experience-transcendent. After all, even if I see the table be painted green, how would things look different if the table were destroyed and replaced by a distinct table like it in many respects except green? Plausibly they wouldn't. If we have justification for claims about essence, they are (in part) a priori.

Accordingly, extreme empiricists think that one is never justified in accepting or rejecting claims of form o is essentially F. Accordingly, extreme empiricists should think there's not much point to talking about essentiality, and the rest of the philosophy that goes along with it.

Some amateur history of philosophy

Hume was more permissive than an extreme empiricist. Hume thought that one is (noninferentially) justified in judging that p iff (i) one's perceptual experience has p as part of its content (ii) p *analytic*, the content of some representation which is analytic, which one is justified in accepting purely in virtue of one's *semantic competence*, or one's grasp of logic and definitions. Hume bought into the *analytic theory of the a priori*.

Kant was more permissive than Hume. Kant thought that one is (noninferentially) justified in judging that p iff (i) one's perceptual experience has p as part of its content (ii) p is analytic (iii) p is *synthetic a priori*, the content of some representation which is fundamental to mental structures which constrain one's perceptual and cognitive experiences.

Kant added category (iii) because he did not see how Hume could account for our knowledge of mathematics.

The *logical empiricists*—Carnap included—bought the analytic theory of the a priori. They thought that Kant's argument against Hume went wrong in that Kant and Hume were ignorant about the nature of logic. They used antiquated Aristotelian logic rather than fancy Fregean logic. They accepted the *logicist* thesis that, in Fregean logic, one is justified in accepting the axioms of mathematics purely on the basis of one's semantic competence (how this is supposed to be done, and whether it works, needn't detain us).

What Carnap is up to

Carnap's primary philosophical aim is to show how claims of essence can be treated as analytic;

to do this, he develops a semantical theory which is very tightly linked to the pattern of possibilities.

Does he succeed?

As we will see, this gets into very deep questions.

2 L-truth

State-descriptions

Fundamental to the project is the notion of a *state-description*. This is supposed to correlate to the notion of a *possible world* in Leibniz's system. However, while possible worlds are supposed to be *worlds*, made up of (perhaps merely possible) matter and energy, with pink and shiny (merely possible?) things as parts, Carnap's state-descriptions are *linguistic* entities, containing only words and more complex linguistic expressions as parts.

Suppose we have a language with three predicates 'F', 'G', and 'H', and two names 'a' and 'b'. Then the language contains six atomic sentences: 'Fa', 'Fb', 'Ga', 'Gb', 'Ha', and 'Hb'.

A state description is a list of sentences which contains, for each atomic sentence, either it or its negation. For instance:

- Fa, Fb, Ga, Gb, Ha, Hb;
- $Fa, \neg Fb, Ga, \neg Gb, Ha, \neg Hb$;
- $\neg Fa, \neg Fb, Ga, \neg Gb, Ha, \neg Hb$.

For our mini language, there will be $2^6 = 64$ state descriptions (why?).

> A class of sentences in S_1, which contains for every atomic sentence either this sentence or its negation, but not both, and no other sentences, is called a ***state-description*** in S_1, because it obviously gives a complete description of a possible state of the universe of individuals with respect to all properties and relations expressed by predicates of the system. Thus the state-descriptions represent

> Leibniz's possible worlds or Wittgenstein's possible states of affairs. (9)

Of the state-descriptions, one of them is *actual*, the one all members of which are *true*. The others correspond to *merely possible worlds*.

Truth in a state-description

Which sentences are true in a state-description? Clearly the ones in the list. But more than just that: in the first one, we want to say that '$\neg\neg Fa$', '$Fa \wedge Fb$', and '$\exists x Fx$' are true.

The rules for the first two are straightforward:

- If $\ulcorner \neg S \urcorner$ is a sentence of the language, it is true in a state-description iff S is not true in the state description;
- if $\ulcorner S \wedge S' \urcorner$ is a sentence of the language, it is true in a state-description iff S is true in the state-description and S' is true in the state-description.

For quantification, Carnap gives a "substitutional" interpretation:

A $\ulcorner \forall x \varphi(x) \urcorner$ is true in a state-description iff for every name n in the language, $\ulcorner \varphi(n) \urcorner$ is true in the state-description (9(5)).

E Existential quantification is defined as the "dual" of universal quantification, or, equivalently:

$\ulcorner \exists x \varphi(x) \urcorner$ is true in a state-description iff for some name n in the language, $\ulcorner \varphi(n) \urcorner$ is true in the state-description.

Substitutional quantification is a bit weird. Consider rule (A): it says that what we mean by the claim that *everything* is F is that *everything we (in principle?) have a name for* is F. That's not right unless we think we can in principle name anything. (Maybe this is a theory of the meanings of restricted quantifiers: everything *we care about* is F, and we don't care about anything we're not in principle in a position to name?)

Meaning and truth in a state description

The sorts of rules just given state facts about meaning, as to know what S means is (in part?) to know its truth-condition:

> The class of all those state-descriptions in which a given sentence $\mathfrak{S}_i$ holds is called the ***range*** of $\mathfrak{S}_i$. All the rules together [...] determine the range of any sentence in S_1; therefore, they are called

> *rules of ranges*. By determining the ranges, they give, together with the rules of designation for the predicates and the individual constants, an *interpretation* for all sentences in S_1, since to know the meaning of a sentence is to know in which of the possible cases it would be true and in which not, as Wittgenstein has pointed out. (9–10)

L-truth

L-truth abbreviates "logical truth"; it is supposed to be an "explication" of—a sort of "rational reconstruction" of, or mathematically precise nearby replacement for—the intuitive or philosophical notion of analyticity:

> [Necessary or analytic truth] has sometimes been characterized as truth based on purely logical reasons, on meaning alone, independent of the contingency of facts. Now the meaning of a sentence, its interpretation, is determined by the semantical rules [. . .]. Therefore, it seems well in accord with the traditional concept [. . .] if we require of any explicatum that it fulfil the following condition:
>
> 2-1. *Convention*. A sentence $\mathfrak{S}_i$ is ***L-true*** in a semantical system *S* iff $\mathfrak{S}_i$ is true in *S* in such a way that its truth can be established on the basis of the semantical rules of the system *S* alone, without any reference to (extra-linguistic) facts. (10)

Note that if 'L-truth' can be defined so that this convention is met, then an L-true sentence will therefore be analytic, as merely understanding it (perhaps together with a bit of calculation) will suffice for being justified in accepting it.

The proposed definition is that *S* is L-true $=_{\text{df}}$ *S* is true in every state-description:

> A way [to define L-truth] is suggested by Leibniz's conception that a necessary truth must hold in all possible worlds. Since our state-descriptions represent the possible worlds, this means that a sentence is logically true if it holds in all state-descriptions. This leads to the following definition:
>
> 2-2. *Definition*. A sentence $\mathfrak{S}_i$ is ***L-true*** (in S_1) $=_{\text{df}}$ $\mathfrak{S}_i$ holds in every state-description (in S_1).

Here's the argument that definition 2-2 satisfies convention 2-1:

> The following consideration shows that the concept of L-truth thus defined is in accord with the convention 2-1 and hence is an adequate explicatum for logical truth. If $\mathfrak{S}_i$ holds in every state-description, then the semantical rules of ranges suffice for establishing this result. [...] Therefore, the semantical rules establish also the truth of $\mathfrak{S}_i$ because, if $\mathfrak{S}_i$ holds in every state-description, then it holds also in the true state-description and hence is itself true. If, on the other hand, $\mathfrak{S}_i$ does not hold in every state-description, then there is at least one state-description in which $\mathfrak{S}_i$ does not hold. If this state-description were the true one, $\mathfrak{S}_i$ would be false. Whether this state-description is true or not depends upon the facts of the universe. Therefore, in this case, even if $\mathfrak{S}_i$ is true, it is not possible to establish its truth without reference to facts. (11)

This is a very cool and persuasive argument: we will come back to it next week.

3 Intension and extension

Carnap gives a two-factor account of meaning on the basis of a Fregean referential semantics together with the notion of L-truth.

Extension

For two expressions of the same category (two terms, two predicates, two sentences), say they *have the same extension* iff:

- Terms: t and t' have the same extension iff $\ulcorner t = t' \urcorner$ is true;
- Predicates (unary): F and F' have the same extension iff $\ulcorner \forall x(Fx \text{ iff } F'x) \urcorner$ is true;
- Sentences: S and S' have the same extension iff $\ulcorner S \text{ iff } S' \urcorner$ is true.

We can identify what the extension of an expression should be, such that these rules come out:

- The extension of a *term* is its (actual) referent;
- the extension of a *predicate* is the "class" or set of entities to which it actually applies;
- the extension of a *sentence* is its truth-value.

This looks an awful lot like Frege.

Note that extension is tied to actuality.

Recall the concern Frege addressed in ‘Function and Concept’, about the view that the referent of a sentence is its truth-value, that on a purely referential semantics this means all true sentences mean the same thing. Frege adopts a two-factor view (reference + sense) to block this concern.

A similar concern arises on a purely extensional version of Carnap’s view. If two expressions had the same meaning iff they have the same extension:

- All coreferential terms would have the same meaning;
- all predicates applying to the same things would have the same meaning (suppose that exactly the same things have hearts and kidneys: then ‘has a heart’ and ‘has a kidney’ would have the same meaning);
- all sentences with the same truth value would have the same meaning.

To block this concern, Carnap adds a second component of meaning.

Intension

For two expressions of the same category (two terms, two predicates, two sentences), say they have the same intension under the following circumstances:

- Terms: t and t' have the same intension iff $\ulcorner t = t' \urcorner$ is L-true;
- Predicates (unary): F and F' have the same intension iff $\ulcorner \forall x(Fx \text{ iff } F'x) \urcorner$ is L-true;
- Sentences: S and S' have the same intension iff $\ulcorner S \text{ iff } S' \urcorner$ is L-true.

Two expressions e and e have the same intension $=_{\text{df}}$ $\ulcorner e \equiv e' \urcorner$ is true.

Adding intension as part of meaning blocks the bad consequences. Here’s how it works for terms:

> Suppose ‘$h = p$’ is not analytic. Then there is some state description in which ‘$h = p$’ is false; perhaps because there is some predicate F such that $\ulcorner Fh \urcorner$ and $\ulcorner \neg Fp \urcorner$ are part of that state-description. Accordingly, ‘h’ and ‘p’ have different intensions, and therefore have different meanings.

Note that for all terms which are not analytic equivalents, identity statements between them will not be L-true.

What are intensions?

Here's Carnap's view:

- Predicates: the intension of a predicate is a *property*—*having a kidney*, *having a heart*, plausibly different things.
- Sentences: the intension of a sentence is a *proposition*—the proposition that Toronto is in Ontario, the proposition that Chicago is in Illinois, plausibly different things.
- Terms: the intension of a term is an *individual concept*. What's an individual concept? Allegedly, something a bit like a property but "of individual type". Not clear I really grok this—more later.

On a more contemporary view, intensions are functions from worlds to extensions:

- Terms: the intension of a term t is a function which takes w into the entity x such that at w, x is the extension of t;
- Predicates: the intension of a predicate F is a function which takes w into the set of things such that at w, those things are the extension of F;
- Sentences: the intension of a sentence S is a function which takes w into the truth-value such that at w, that truth-value is the extension of S.

Moral: the semantics provides a close tie between meaning and modality.

4 Quantified modal logic

Throughout I'll replace Carnap's 'N' with the contemporary '$\Box$', intended to abbreviate 'it is necessary that'.

Let's now show how Carnap provides rules of ranges for sentences of form

Q $\exists x \Box Fx$;

sentences which "quantify in" to the scope of a modal operator. Sentences like this say in effect that something is essentially F.

Contrast (Q) with '$\Box\exists xFx$', which says that necessarily something is F. This can be true even if (Q) is false and nothing is essentially F. This could be true if the thing which is F varies from world to world; not so for (Q).

Propositional modal logic

Carnap gives a "convention" and a provisional definition that is functional for propositional modal logic. The convention is that $\ulcorner \Box S \urcorner$ is to hold iff S is L-true. The provisional definition is that $\ulcorner \Box S \urcorner$ is true in every state-description iff S is true in every state-description.

> 39-1. For any sentence '...', '$\Box(\ldots)$' is true iff '...' is L-true.
>
> We shall construct the system S_2 by adding to the system S_1 the sign '$\Box$' with suitable rules such that the convention just stated is fulfilled. [...]
>
> If we had only sentences without variables, we could simply take the rules of ranges for S_1 [...] and add the following rule:
>
> 41-1. $\Box(\mathfrak{S}_i)$ holds in every state-description if $\mathfrak{S}_i$ holds in every state-description; otherwise, $\Box(\mathfrak{S}_i)$ holds in no state-description.

Note now that facts about the distribution of modals across sentences can be known straightforwardly thanks to our semantic competence:

> Suppose that 'L-true in S_2' is defined in such a way that our earlier convention 2-1 [...] is fulfilled. Let 'A' be an abbreviation for an L-true sentence in S_2 (for example, '$Hs \vee \neg Hs$'). Then '$\Box$(A)' is true, according to 39-1. And, moreover, it is L-true, because its truth is established by the semantical rules which determine the truth and thereby the L-truth of 'A', together with the semantical rule for '$\Box$', say 39-1. (174)

Rules of ranges for open sentences

Open sentences don't get true/false in a state-description "absolutely", but only *relative to an assignment*, or relative to particular substitutions of constants for the variables in the sentence (41-2).

Here's the rule for atomic sentences with one free variable:

> $\ulcorner Fx \urcorner$ is true in a state-description relative to an assignment of n to 'x' iff $\ulcorner Fn \urcorner$ is true in the state-description. (41-2a)

He gives other rules but these need not detain us. (Technical note: it's not obvious these are necessary, as the substitutional theory of quantification ensures that when it comes time to evaluate the bits in the scope of quantifiers, free variables are never encountered.)

Quantifying in

Let's now analyze (Q).

By (E), (Q) is true in a state description iff for some name n, $\ulcorner\Box Fn\urcorner$ is true in that state description.

By (41-1), (Q) is true in a state description iff for some name n, $\ulcorner Fn\urcorner$ is true in every state description. (Since the RHS doesn't mention the state-description on the LHS, (Q) is true in *every* state description iff for some name n, $\ulcorner Fn\urcorner$ is true in every state description.)

Quantifying in and names

Suppose we want to analyze (Q1) 'Bill is such that *he* is essentially F'. The relevant logical form would be

$$\exists x(x = b \wedge \Box Fx).$$

And the relevant Carnapian analysis would be the following:

* (Q1) is true iff:
for some name n:
$\ulcorner n = b\urcorner$ is true and $\ulcorner Fn\urcorner$ is true in every state description.

Carnapian QML and the epistemology of essence

This can start to look good for the Humean epistemologist of essence. Suppose that part of what 'Bill' means to me is '42d President of the US'. Then if F = '42d President of the US', there *is* some name n such that the above condition can be known to be met merely on the basis of my semantic competence, namely 'Bill'.

So it looks like certain facts about essence have been shown to be analytic.

5 A Quinean dilemma

Suppose that 'p' abbreviates 'Phosphorus' and 'h' abbreviates 'Hesperus', and 'M' abbreviates 'is visible in the morning'. Suppose we want 'Mp' to be analytic but not so for 'Mh': then some state descriptions contain '$\neg Mh$', others contain 'Mh', but all state descriptions contain 'Mp'. We know that '$h = p$' is true in the actual state description.

Then there's a dilemma. Either:

- Essence outstrips rules of language; or

- rules of language outstrip essence; or
- it doesn't make sense to speak of the absolute essence of a thing; or
- we abandon the analytic theory of essence entirely.

This is something like the concern Quine raises in 'Reference and necessity' (though Quine overlooks the third option).

Too many essential properties?

Now consider the analysis of 'Phosphorus is such that *it* is essentially visible in the morning'. Along the lines of (*), this will be true iff:

- for some name n:
  $\ulcorner n = p \urcorner$ is true and $\ulcorner Mn \urcorner$ is true in every state description.

There is such an n: namely, 'p'. Hence, the sentence comes out true. So far so good.

But now consider the analysis of 'Hesperus is such that *it* is essentially visible in the morning'. Along the lines of (*), this will be true iff:

- for some name n:
  $\ulcorner n = h \urcorner$ is true and $\ulcorner Mn \urcorner$ is true in every state description.

Once again, there is such an n: namely, 'p'. Hence, the sentence comes out true.

And that might seem very bad for Carnap. After all, it's *not* knowable in virtue of meaning that Hesperus is essentially visible in the morning! At best, we know this because we know in virtue of meaning that *Phosphorus* is essentially visible in the morning, and we know that Hesperus and Phosphorus are identical.

Too few essential properties?

Suppose we instead desire the truth-conditions to be as follows:

- for every name n such that $\ulcorner n = p \urcorner$ is true,
  $\ulcorner Mn \urcorner$ is true in every state description;
- for every name n such that $\ulcorner n = h \urcorner$ is true,
  $\ulcorner Mn \urcorner$ is true in every state description.

Clearly to get either of these results, we would have to fiddle around with (E), perhaps along the following lines:

E′ $\ulcorner \exists x \varphi(x) \urcorner$ is true in a state-description iff for some name n in the language, for every name m such that $\ulcorner m = n \urcorner$ is true, $\ulcorner \varphi(m) \urcorner$ is true in the state-description.

Under this regime, 'Hesperus is such that it is essentially visible in the morning' is false, desirably. But, also undesirably, 'Phosphorus is such that it is essentially visible in the morning' is also false.

Essences of "individual concepts"?

Maybe the logical form of (Q1) is not

$$\exists x(x = b \wedge \Box Fx),$$

but rather

$$\exists x(x \equiv b \wedge \Box Fx).$$

Then the relevant Carnapian analysis would be the following:

** (Q1) is true (in the actual state-description) iff:
for some name n:
$\ulcorner n \equiv b \urcorner$ is true and $\ulcorner Fn \urcorner$ is true in every state description.

Recall that '$s \equiv t$' means that the *individual concept* s is identical to the *individual concept* t.

Note that the Phosphorus-individual concept and the Hesperus-individual concept are distinct entities: there is some state description in which 'Mp' is true but 'Mh' is false. Leibniz's law establishes distinctness of objects when the violations are at the *actual* state-description, but distinctness of individual concepts when the violations are at nonactual state-descriptions.

Accordingly, when we agree to both 'Phosphorus is essentially visible in the morning' and 'Hesperus isn't essentially visible in the morning' we are talking about individual concepts and not about individuals.

This seems to be Carnap's actual view (section 44).

What's an individual concept?

> Let us look for entities which we might regard as intensions of individual expressions. According to our definition for the identity of intensions (5-2), the intension must be something that L-equivalent individual expressions [...] have in common. We have earlier found entities which seemed suitable as intensions of designators of other types [...]. In these cases, the intensions are those

> entities which are sometimes regarded as the meanings of the expressions in question; and, in the case of predicators and functors, the intensions are concepts of certain types. Now it seems to me a natural procedure, in the case of individual expressions, likewise to speak of concepts, but of concepts of a particular type, namely, the individual type. Although it is not altogether customary to speak here of concepts in this sense, still it does not seem to deviate too much from ordinary usage. I propose to use the term '***individual concept***' for this type of concept. (40–1)

My sense of this is that I don't get what an individual concept is, and for that reason I doubt that (**) is the right analysis of (Q1). I thought that when I said 'Bill is such that *he* is essentially *F*', I was talking about *Bill*, and not about some other entity, the "Bill-individual concept". This is in essence Quine's complaint in section 44.

Some of our contemporaries follow David Lewis and believe that things have essences not absolutely, but relative to a way of thinking or speaking about them. The thought is that since it is analytic that Phosphorus is visible in the morning but not that Hesperus is visible in the morning, then the following are true:

- Phosphorus is such that *it* is essentially visible in the morning *as Phosphorus*;
- Hesperus is not such that *it* is essentially visible in the morning *as Hesperus*.

Here the thought is something like this: if we compare something in the actual world—Hesperus—some other possible world, there may be no absolute answer to the question, which thing in that other possible world is identical to Hesperus. Rather, there may be many candidate things which are identical to Hesperus relative to different ways of "individuating" objects.

Perhaps we can think of the individual concept Hesperus as that sum or set of merely possible objects which are identical to the planet Venus *relative to the Hesperus-y style of individuation*; the individual concept Phosphorus as that sum or set of merely possible objects which are identical to the planet Venus *relative to the Phosphorus-y style of individuation*. All of the latter but not all of the former are visible in the morning.

"Aristotelian essentialism"?

If one dislikes relativism about essence, and wishes to avoid both overpopulating essences (by associating *x*'s essential properties with analytic entailments of *some* name for *x*) and underpopulating essences (by associating *x*'s essential properties with analytic entailments of *every* name for x), there may be no choice but to untie the theory of essence from analyticity entirely. In that case we would either need some new epistemology of the a priori, or else show that the epistemology of essence is not entirely a priori.