Radical Interpretation, Part II: Philosophical Logic.

Lecture IV, *David Lewis on Radical Interpretation*, 27th November.

Christopher J. Masterman (cm789@cam.ac.uk, christophermasterman.com)

Last week, we looked at Davidson on radical interpretation. There were two key takeaways.

1. Davidson does indeed think that radical interpretation can be done—it is possible, given the relevant physical facts, to formulate interpretive theories of meaning for completely unknown languages, i.e., Tarski-style theories of truth for such unknown languages, which assign meanings to sentences.
2. There remains, for Davidson, an indeterminacy at the sub-sentential level. In particular, there is indeterminacy of reference. This is supposed to be unsurprising: Davidsonian theories of meaning are naturalist theories of public, linguistic phenomena. Notions like *reference* behave as theoretical posits.

This week, we'll look at David Lewis's discussion of radical interpretation in (Lewis, 1974).

1. *David Lewis's Radical Interpretation*

1.1. Both Quine's and Davidson's radical translation/interpretation are first and foremost concerned with sentences: Quine in pairing sentences of different languages *via* translation manuals and Davidson in interpreting an unknown language—developing interpretive theories of meaning which fit the physical data, where this data includes facts about the informant *holding true* certain sentences. David Lewis contrasts here in that he takes radical interpretation to have both a broader *aim* and tied to a broader *set of data*.

(*Note that Lewis, like Davidson, focuses his attention on some individual 'Karl'*)

Broader Aim: The task is 'of coming to know Karl as a person.' (Lewis, 1974: 331) That is:

> We would like to know what he believes, what he desires, what he means, and anything else about him that can be explained in terms of these things. We seek a two-fold interpretation: of Karl's language, and of Karl himself. (Lewis, 1974: 331)

Broader Set of Data: We are to carry out this task of coming to know Karl as a person, with limited knowledge—'limited to our knowledge of him as a physical system' (331). But this includes '... the whole truth about Karl as a physical system':

> '...how Karl moves, what forces he exerts on his surroundings...'
>
> '...what light or sound or chemical substances he absorbs or emits...'
>
> '...the same things about all of Karl's material parts, great or small, permanent or temporary.'
>
> '...all the masses and charges of the particles that compose him...'
>
> '...all the magnitudes and directions of the fields and potentials and radiation that pervade him...'
>
> '...not only his present physical state but also his physical history...'
>
> '...also the nomic or counterfactual or causal dependencies among [actual particular physical facts]...' *etc.*

1.2. Radical interpretation, for Lewis, is the task of understanding a person in this very broad sense from the totality of all physical facts. Lewis frames his discussion of the task of radical interpretation in terms of the relations that hold between four different classes of facts: **P**, **Ao**, **Ak**, and **M**:

P: The whole truth of Karl as a physical system.

Ao: Karl's attitudes, beliefs and desires, as expressed in *our* language.

Ak: Karl's attitudes, beliefs and desires, as expressed in *Karl's* language.

M: Karl's meanings: truth conditions of his full sentences, including the denotations. (Lewis, 1974: 332).

1.3. We have discussed **P**. **Ao** and **Ak** are specifications of Karl's propositional attitudes. These specifications are done in a particular language. In **Ao**, Karl's propositional attitudes are specified in our language, whereas in **Ak** Karl's propositional attitudes are specified in Karl's language. Specifications have the form:

$$\text{Karl}\left(\frac{\text{believes}}{\text{desires}}\right)\text{to degree } d\text{, at } t\text{, the proposition expressed in } c \text{ by the sentence } S \text{ of}\left(\frac{\text{our}}{\text{Karl's}}\right)\text{language.}$$

M is the specification, in our own language, of the meaning of expressions of Karl's language. **M** will contain specifications of the truth conditions of full sentences of Karl's language, as well as the syntactic and semantic rules which are capable of generating sentences of Karl's language and their truth conditions.

1.4. Radical interpretation is the task of showing how **P** determines all of the other facts, i.e., **Ao**, **Ak**, and **M**. In Lewis's phrase, we are given **P**, and we must solve for the rest. Crucially, the task here is not epistemological—it is not how we come to know **Ao**, **Ak**, and **M**. We are idealising here, by holding fixed all of **P** and investigating how they determine **Ao**, **Ak**, and **M**. (Note how this differs from Davidson's radical interpretation, which is plausibly understood as asking how **Ak** determines **M**.)

2. *How Lewis Radically Interprets*

2.1. Lewis takes it that the problem of radical interpretation—given **P**, solve for **Ao**, **Ak**, and **M**—is solved since the task of radical interpretation is constrained enough by certain fundamental principles. These constraints are principles from *our fundamental theory of persons*. These constraints are partly definitional in that they tell us what it is to have a belief, a desire, etc.—the less someone conforms to such constraints, the less of a claim can be made that they have the relevant mental state. Although they may appear platitudinous, the thought is that they are central to what it is to have, e.g., a belief, and so in solving for **Ao**, **Ak**, and **M** from **P**, we should be constrained by such principles. (Think back to Davidson on Principle of Charity.)

2.2. Lewis proposes six constraints, *Charity*, *Rationalisation*, *Truthfulness*, *Manifestation*, *Generativity*, and *Triangle*. The first four—*Charity*, *Rationalisation*, *Truthfulness*, *Manifestation*—can be thought to be partly definitive of propositional mental states like belief and desire. That is, they place constraints on the relationships that ought to hold between **P**, **Ao**, and **Ak**. In more detail:

Principle of Charity: This constrains the relationship between **Ao** and **P**: 'Karl should be represented as believing what he ought to believe, and desiring what he ought to desire' (Lewis, 1974: 336). That is, how *we* represent Karl's propositional attitudes in *our* language (**Ao**) should be charitable in this way.

- How so, according to Lewis? We should attribute to Karl propositional attitudes which we would have, *were* we in Karl's position. So, given the same physical inputs as you, as dictated in **P**, Karl will form the same beliefs as you. If there are differences between Karl's position and our own, again as dictated by **P**, then we should attribute the beliefs we *would* have (Lewis, 1974: 336–7)

Principle of Rationalisation: This constrains the relationship between **Ao** and **P**: 'Karl should be represented as a rational agent; the beliefs and desires ascribed to him by **Ao** should be such as to provide good reasons for his behavior, as given in physical terms by **P**.'

- Our interpretation of Karl's propositional attitudes should allow us to subsume his actions, coupled with his propositional attitudes, specified by **Ao**, under an adequate decision theory.

Principle of Truthfulness: This constrains the relationship between **Ao**, **M**, and **P**. Generally, we should take Karl to be governed by the usual constraint of asserting or believing that which is true and being governed by the constraint that generally communication is an effort to communicate such truth.

The Manifestation Principle: This constrains the relationship between **P** and **Ak** (and thus **P** and **Ao**, if we are going to correctly interest Karl's propositional attitudes): 'Karl's beliefs, as expressed in his own language, should normally be manifest in his dispositions to speech behavior' (Lewis, 1974: 339)

- Note that *normally* is important here: Karl may intend to deceive and thus his disposition to utter S, as determined in **P**, does not mean that **Ak** should contain belief in the proposition expressed by S. **Ak** should normally just straightforwardly track dispositional behaviour as determined by **P**.

2.3. The remaining constraints do not concern how **Ao**, **Ak**, and **M** relate to **P**. Rather, they are in a sense internal, constraining only **Ao**, **Ak**, and **M** alone with no direct input from **P**.

The Principle of Generativity: This constrains **M**: '**M** should assign truth conditions ... in a way that is at least finitely specifiable, and preferably also reasonably uniform and simple.' (Lewis, 1974: 339).

- Lewis is open about which form the syntactic/semantic rules should take. (Recall that Davidson secures Generativity with theories of meaning as finitely axiomatizable Tarski-style theories of truth.)
- Why Generativity? Lewis doesn't say, but we can take him to be swayed by issues of learnability: Karl cannot plausibly know his own language if **M** is *not* finitely specifiable.

The Triangle Principle: This constrains how **Ao**, **Ak**, and **M** interact: 'Karl's beliefs and desires should be the same whether expressed in his language or in ours' (Lewis, 1974: 339). This plays out as follows.

> Suppose that **M** assigns a certain truth condition to a sentence s ... in Karl's language, and suppose that a sentence s' ... of our language has the same truth condition. Then if **M** is correct ... we ought to be entitled to regard s ... and s' ... as expressing the same propositions in their respective languages. If so ... if **Ak** ascribes to Karl a certain degree of belief in the proposition expressed by s ..., then **Ao** should ascribe to him the same degree of belief in the proposition expressed by s' ... and likewise for Karl's degrees of desire. (Lewis, 1974: 339)

- Think of the Triangle Principle as a constitutive constraint on what it is the be the content—or proposition—of a propositional attitude, as opposed to a definitional constraint on mental states.

3. *Lewis's Solutions to the Problem of Radical Interpretation.*

3.1. At the end of (Lewis, 1974), Lewis discusses three methods for solving the problem of radical interpretation. The first method sticks closely to how Davidson proposes to radically interpret in (Davidson, 1973). Lewis argues that the Davidsonian approach is wanting for a couple of reasons. The chief reason is that it places too much emphasis on *language* as a vehicle for the manifestation of belief and desires. That is, in Davidson's solution to radical interpretation, little role is played by *Rationalisation*—little focus is placed on language as a social enterprise and little focus is placed on how belief manifests in non-linguistic behaviour.

3.2. Lewis doesn't think that the Davidsonian approach is bound to fail to give a determinate solution to the problem of radical interpretation. Perhaps, he conjectures, it would work if we made better use of the other constraints. But this is by the by. Lewis's preferred method can be detailed in the following steps.

Method Two (Lewis's Preferred Method)

Step One. By appealing to *Rationalisation* and *Charity*, fill in **Ao** with reference to the facts in **P**.

(For Lewis, we fill in **Ao** non-tentatively. *Rationalisation* and *Charity*, given **P**, give us full specification of **Ao**.)

Step Two. Using **Ao**, fill in **M** folllowing *Generativity* and *Truthfulness*.

Step Three. Given **Ao** and **M**, fill in **Ak** using *Triangle*. The solution is now complete.

(Lewis notes that *Manifestation* is automatically satisfied. That is, *Manifestation* is redundant on this method.)

3.3. It's important to stress how we are supposed to take Lewis's method for solving radical interpretation here. One may of course worry that we cannot simply assume that each of the relevant steps can be successfully carried out. Moreover, Lewis provides little argument that any of these can be done successfully. The point, rather, is to detail the general shape of how a solution could be found and which constraints any solution should be beholden to. Earlier in (Lewis, 1974), he writes:

> If I ask how **P** determines all the rest, my question requires the presupposition that **P** does determine all the rest. Or, at least, that **P** determines all the rest to the extent that anything does—that where determination by **P** leaves off, there indeterminacy begins (Lewis, 1974: 334)

3.4. So what about indeterminacy? *How* many solutions to the problem of radical interpretation might there be, e.g., uniquely one? equally adequately many? Lewis accepts that there will be at least two forms of indeterminacy. First, Lewis accepts (along with Quine and Davidson) that the truth conditions for full sentences, given in **M**, will not be alone enough to determine the meanings of the constituent parts, even assuming *Generativity*. Second, Lewis accepts that a more general indeterminacy in the specification of **M** could arise if no solution fits the constraints perfectly—cases of involving 'the confused desires of a compulsive thief' (Lewis, 1974: 343). For Lewis, a more serious kind of indeterminacy is one which would arise if there were two incompatible solutions which *fit the constraints perfectly*. However:

> *Credo*: if ever you prove to me that all the constraints we have yet found could permit two perfect solutions, differing otherwise than in the auxiliary apparatus of **M**, then you will have proved that we have not yet found all the constraints (Lewis, 1974: 343)

References

Davidson, Donald (1973). Radical Interpretation. *Dialectica* 27, 313–328.
Lewis, David K. (1974). Radical Interpretation. *Synthese* 23, 331–344.