Radical Interpretation

J. Butterfield and A. Caulton, Thursday 18 November 2010

1 Quine and Davidson: a (very) brief reminder

Cf. Quine (1960, Ch. 2; 1969), Davidson (1973).

(Context) The meanings of any language’s terms (sub-sentential components) is constrained only by the truth-conditions of sentences in which they appear. (Take truth-conditions as maps from indices—i.e. ordered ‘tuples containing a world and context-determinants (time, place, speaker, audience, indicated-objects, previous discourse) . . .—to truth-values.)

(Assent/Dissent) A community C’s actual linguistic behaviour—their actual pattern of assent and dissent to particular utterances—only partially constrains the truth-conditions of the sentences of their language L. (Take L to comprise a vocabulary and a syntax, and their products (i.e. grammatical constructions).)

(Context) and (Assent/Dissent) entail (Underdet):

(Underdet) A community C’s actual linguistic behaviour radically underdetermines L’s meanings.

(Behaviour) C’s actual linguistic behaviour exhausts what can fix meanings in L.

(Underdet) and (Behaviour) entail (Radical):

(Radical) The meanings in L are radically underdetermined.

(Radical) is intended to explode the “myth of the museum in which the exhibits are meaning and the words are labels. To switch the language is to change the labels.” (Quine 1969, p. 27)

. . . [R]ival systems of analytical hypotheses [term-for-term translation manuals] can conform to all speech dispositions within each of the languages concerned and yet dictate, in countless cases, utterly disparate translations; not mere mutual paraphrases, but translations each of which would be excluded by the other system of translation. Two such translations might even be patently contrary in truth value, provided there is no stimulation that would encourage assent to either. (Quine 1960, pp. 73-4)

2 Preliminaries

Lewis (1974) hopes to mitigate (Radical)—or avoid it altogether—by denying (Behaviour) and introducing new constraints on interpretation/translation. Lewis proposes replacing it with

(MinMaterial) The local physical facts P about members of C exhausts what can fix meanings in L.

(‘Physics’ is to be taken much wider than ‘contemporary physics’.)

Note that: (i) Lewis originally relativises meaning to a person’s language, not a community’s, but alters this commitment in the Postscript: (ii) P will include a catalogue of the speakers’ “surface irritations”, and their overt behaviour; (iii) Lewis takes it that P suffices to determine speakers’ mental states: but (iv) P will suffice to determine only (if at all) the narrow content of speakers’ mental states (cf. Oskar and Pierre).
3 Lewis’s radical interpretation project

Let us focus on a particular member of $C$, called Karl. We can categorise all the physical, mental and semantic facts about Karl as follows:

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure.png}
\caption{Figure taken from Lewis (1974), p. 109. All page refs. apply to the reprinted paper, in Lewis (1983).}
\end{figure}

“Given $P$, the facts about Karl [and others] as a physical system, solve for the rest.”

- The project is not the practical project of a field linguist determining the semantic facts, but the metaphysical project to account for how $P$ determines all the mental and semantic facts.

- $A_0$ and $A_k$ must contain not only propositional attitudes, but irreducible “attitudes de se”, of which the propositional attitudes are a part (cf. Lewis, 1979). More than just propositional attitudes may be expressed with sentences, with the use e.g. of personal indexicals (cf. RI, p. 121).

- Lewis leaves it undecided whether $M$ contains an complete specification of truth-conditions, or just a specification for “not too distant” possible worlds, or just a specification of actual truth-values.

The solution:

What are the constraints by which the problem of radical interpretation is to be solved? Roughly speaking, they are the fundamental principles of our general theory of persons. They tell us how beliefs and desires and meanings are normally related to one another, to behavioral output, and to sensory input.

The general theory of persons serves as a schema for particular theories of particular persons. A particular theory of Karl, for instance, may be constructed by ascribing particular beliefs, desires, and meanings to him. That is: by filling in $A_0$, $A_k$, and $M$. But not just any filling-in will do. The relations of $A_0$, $A_k$, and $M$ to one another and to $P$ must conform—for the most part, more or less—to the principles of the general theory. (RI, p. 111)

The general theory of persons is exactly the one found in Lewis (1970, 1972). We “particularise” the general theory for Karl by filling in Karl’s specific surface irritations, neural states, overt behaviour, etc., as given by $P$. Karl’s particular beliefs, desires, and meanings are then the (hopefully) unique realizers of their appropriate roles, as dictated by the general theory.
The details (I): six constraining principles (perhaps redundant, perhaps incomplete)

1. *Charity.* (Constrains \( A_0 \), given “surface irritations” in \( P \).) This does not mean filling \( A_0 \) with *our* beliefs and desires, it means filling \( A_0 \) with what *we would* believe or desire if we were in Karl’s place. (E.g. past scientists.) (“But that’s only our opinion! Yes. Better we should go by an opinion we don’t hold?”)

So we must find some common inductive method \( M \) which takes us from evidence to beliefs (and takes us from *our* evidence to *our* beliefs; and some “common underlying system of basic intrinsic values” \( \mathcal{U} \), which takes us from beliefs to desires (and takes us from *our* beliefs to *our* desires):

\[
\begin{array}{ccc}
\text{Karl's evidence} & \rightarrow & \text{Karl's beliefs} \\
\downarrow & & \downarrow & & \downarrow & & \downarrow \\
\mathcal{M} & & \text{our evidence} & \rightarrow & \text{our beliefs} & \rightarrow & \text{our desires} \\
\text{Karl's desires} & \rightarrow & \mathcal{U} & & \text{beliefs} & & \text{desires} \\
\end{array}
\]

(Figure taken from Lewis (1974), p. 113.)

2. *Rationalization.* (Constrains \( A_0 \), given overt behaviour in \( P \).) Karl should be made out to be a rational (or mostly rational) agent: he should be made out to have good reasons for his behaviour. In decision-theoretic terms: \( A_0 \) should be such that Karl’s overt behaviour has the maximum expected utility, where the basic utilities are provided by \( \mathcal{U} \) and the probabilities by Karl’s beliefs. (So Lewis takes decision theory to be part of the general theory of persons.)

3. *Truthfulness.* (Constrains the relation between \( A_0 \) and \( M \).) Amongst Karl’s beliefs and desires must be attitudes which constitute his part in a *convention of truthfulness in* \( L \), Karl’s language. This involves having the desire to use \( L \) appropriately as a *speaker* (e.g. to describe things truthfully); and, as an *audience*, having the beliefs that issue appropriately from uses of \( L \) by others in \( C \) (e.g. to believe certain utterances). What is ‘appropriate’ depends on what the truth-conditions of \( L \) are (i.e. \( M \)), and on what the facts are perceived to be (which are contained in \( A_0 \)).

(This is not enough: Karl must also have higher-order beliefs: that others in \( C \) will act appropriately, as he does; and that others in \( C \) expect him to act appropriately, as they do; and so on. He must also have the desire to act as \( C \) do, so long as they do, and the belief that others in \( C \) have similar desires, and so on. For more details, see Lewis (1969, 1975).)

4. *Generativity.* (Constrains \( M \).) Truth-conditions should be finitely specifiable (and preferably uniform and simple), as in a Tarski-style truth theory (cf. Davidson (1973)).

5. *Manifestation.* (Constrains \( A_k \) (partly \( A_0 \)), given overt behaviour in \( P \).) Karl should *typically* be taken to express belief in what he genuinely believes, and to express desires for what he genuinely desires. Certain beliefs and desires in \( A_0 \) may provide motivation for *atypical* overrides.

6. *Triangulation.* \( A_0 \) and \( A_k \) should entirely agree, given \( M \).

The details (II): two methods

*Method 1* (cf. Quine, Davidson):

1. Use *Charity* and “surface irritations” in \( P \), tentatively to fill in \( A_0 \).
2. Use Manifestation and overt linguistic behaviour in P, paying heed to Ao for potential overrides, to fill in Ak. (Cf. (Assent/Dissent), above.)

3. Use Triangulation and Generativity to fill in M, given Ao and Ak, and revise beliefs in Ao.

4. Use Rationalization, behaviour in P, and beliefs in Ao, revise desires in Ao. Use Triangulation to fill in desires in Ak.

5. If recent revisions to Ao would have made step 2’s results to come out differently, start again at step 2 and repeat. Otherwise end.

Method 1 over-eggs Manifestation and doesn’t use Truthfulness at all, or Rationalization for belief. This ignores the role of language as a social practice, and belief as determining non-linguistic behaviour.

Method 2 (Lewis’s favourite):

1. Use Charity and Rationalization, and all of P (“surface irritations”, all overt behaviour, neural states even?), to completely fill in Ao.

2. Use Truthfulness, and beliefs and desires pertaining to linguistic behaviour in Ao, to fill in M. Use Generativity to tidy up and complete the filling in of M.

3. Using Triangulation, Ao and M, completely fill in Ak. Manifestation should be automatically satisfied, given Truthfulness, Rationalization applied to linguistic behaviour, and Triangulation.

Method 2 avoids the Quine/Davidson triangulation problem altogether.

“Credo: if ever you prove to me that all the constraints we have yet found could permit two perfect solutions . . . , then you will have proved that we have not yet found all the constraints.” (RI, p. 118)

6 References