

## The Twin Earth Thought Experiment

### Putnam's Argument: First Take

- (1) If linguistic internalism is true, then it is not possible that two speakers are in the same mental (psychological) state, but refer to different objects or facts.
- (2) But there is a conceivable case where precisely this is the case.
- (3) Hence, linguistic internalism is false.

### A More Detailed Second Take

- (4) The meaning (sense, *Sinn*, intension) of words reflects the mental states of the speakers: their beliefs, perceptions, intentions and so on. (Narrow) mental states determine the meanings of words (cf. Reading for 25 February).
- (5) Meaning determines (entails) reference (extension, *Bedeutung*). This means that two words with the same meaning have the same reference, and two words with different references cannot have the same meaning. The meaning of a word determines the truth-conditions for its use.
- (6) If two speakers are duplicates, then their mental states are indiscernible from one another. If (4) is true, the meanings of their words is the same. One speaker is on Earth, the other is on Twin-Earth.
- (7) On Earth, 'water' refers to H<sub>2</sub>O.  
On Twin-Earth, 'water' refers to XYZ.
- (8) Hence, the reference of 'water' differs. The use of the word 'water' has different truth-conditions on these two planets: on Earth, 'water' statements are made true by H<sub>2</sub>O, on Twin-Earth, 'water' statements are made true by XYZ. (Note that *A* and Twin-*A* are ignorant about the reference of each other's word 'water'. But it is clear that *A* could not have any experience of XYZ, and Twin-*A* could not have any experience of H<sub>2</sub>O, because these substances do not exist where they are.)
- (9) But if the reference of 'water' differs, then given (5), the meaning of 'water' differs too.
- (10) Yet, given (6), *ex hypothesi*, the mental states of the two speakers are identical.
- (11) Therefore, (4) is false: the (narrow) mental states of speakers do not determine the meanings of words.
- (12) Therefore, the mental states of speakers do not determine the reference of words either. Internalism (or intention-based semantics) is false.

## Häggqvist's Model<sup>1</sup>

Thought experiments are typically designed to show that a target thesis (T) is false. This is brought out by the following argument schema:

- |  |  |
|--|--|
| <p>(a) <math>\Diamond C</math><br/> <math>T \supset (C \Box \rightarrow W)</math><br/> <math>C \Box \rightarrow \neg W</math><br/> <hr style="width: 20%; margin-left: 0;"/> <math>\neg T</math></p> | <p>In words: A counterfactual scenario is possible. Now, if T is true, then given the thought experiment (C), we would expect W to be true. But through the thought experiment we realise that W is false. So, T is false (too).</p> |
|--|--|

In this schema, W is the central claim, which is claimed by the thought experiment to be false, *yet* it is also a claim that T assumes to be true. Hence a tension. And given the counterfactual entailment *and* the possibility of the scenario, it follows that the theory is false.

The model applied to the Twin Earth case:

- (b) Possible: we have twin earth *doppelgängers*.  
 If psychology determines reference/extension (see Handout 6), then if we had *doppelgängers*, they would refer to water with the word 'water'. (Why? Because what 'water' refers to is fixed by the word's intension.)  
 But if we had *doppelgängers*, they would not refer to water with 'water': they would refer to XYZ.  
 So, psychology does *not* determine reference/extension: internalism is false.

The model can also show three possible responses to a thought experiment:

- |   |  |   |
|---|--|---|
| <p>(c) T<br/> <math>\Diamond C</math><br/> <math>T \supset (C \Box \rightarrow W)</math><br/> <hr style="width: 20%; margin-left: 0;"/> <math>\Diamond C \ \&amp; \ (C \Box \rightarrow W)</math></p> | <p>(d) T<br/> <math>\Diamond C</math><br/> <math>C \Box \rightarrow \neg W</math><br/> <hr style="width: 20%; margin-left: 0;"/> <math>\neg(T \supset (C \Box \rightarrow W))</math></p> | <p>(e) T<br/> <math>T \supset (C \Box \rightarrow W)</math><br/> <math>C \Box \rightarrow \neg W</math><br/> <hr style="width: 20%; margin-left: 0;"/> <math>\neg \Diamond C</math></p> |
|---|--|---|

In (c), one defends T by biting the bullet: there is nothing odd about holding W if C was the case, hence T stands. (Perhaps C is just odd.)

In (d), one defends T by denying that T is committed to W's truth in C. (Perhaps C is merely hypothetical and hence shows nothing about W and T.) For Häggqvist, this is the 'irrelevance defence'.

In (e), one defends T by showing that C is impossible, perhaps because C is inconceivable. This 'impossibility defence' resembles a *reductio ad absurdum*: if C leads to two contradictory propositions, then C is false.

<sup>1</sup> Häggqvist, S. (2009). A Model for Thought Experiments. *Canadian Journal of Philosophy*, 39, 55–76.

