

## Argument for theory theory

The theory theorists need to show that we cannot make judgements about the mental states of others without relying on a theory about the role of mental states.

One strategy focuses on the concepts used in these judgments. If applying the concepts used in making judgements about the mental states of others relies on a theory about the role of mental states, then those judgments themselves rely on such a theory.

David Lewis (1972) suggests that meaning of theoretical terms (T-terms) is determined by how they figure in the theory in which they are used alongside observational terms (O-terms):

Suppose we have a new theory, *T*, introducing the new terms  $t_1 \dots t_n$ . These are our T-terms. (Let them be names.) Every other term in our vocabulary, therefore, is an O-term. The theory *T* is presented in a sentence called the *postulate* of *T*. Assume this is a single sentence, perhaps a long conjunction. It says of the entities—states, magnitudes, species, or whatever—named by the T-terms that they occupy certain *causal roles*; that they stand in specified causal (and other) relations to entities named by O-terms, and to one another. (Lewis 1972: 253)

Mental state terms are not observational. And at least on a functionalist conception of the mind, mental states are defined by their causal roles. So the only way of making sense of mental concepts is by conceiving them as theoretical terms. If we follow Lewis, we must assume that understanding the meaning of a mental concept requires grasp of some theory about their role. So when we make judgements about the mental states of others, we rely on a theory about the role of mental states.

## Arguments against theory theory

If we make judgments about the mental states of others through applying a theory, then we will locate the particular case among a range of possible ones. This means that our judgments are an inference to the best explanation: this case *x* most closely fits this theoretical possibility *P*. However, “our attitude to the mental states of others isn’t tentative and indefinitely revisable in the way that our attitude towards a scientific theory usually is” (Carruthers 2004, 261). We can perhaps overcome this objection by adopting a nativist version of theory theory.

Jane Heal (1996) presents a different objection: as an explanation of our ability to make judgements about the mental states of others, the theory theory has to assume that we have a (tacit) grasp of a systematic and general theory of *relevance*.

The theory-theorist is committed to the claim that we have—tacitly at least—solved an extremely important precursor problem to the famous Frame Problem in Artificial Intelligence, namely the problem of providing a general theory of relevance. And this claim is highly implausible.

The Frame Problem here is understood as a general epistemological problem for any kind of computation-based intelligence. As Fodor puts it: “How ... does the machine's program determine which beliefs the robot ought to re-evaluate given that it has embarked upon some or other course of action?” (1983, 114)

As we stressed earlier, we can cope when circumstances are not normal and we understand very well that others can do so too. So if our imagined psychological theory is to account for our competence in these cases it must give systematically organised insight into the difference between our responses in usual and unusual cases, i.e. insight into a whole range of world view/question pairs and their possible upshots. It must specify the range of psychological factors which influence thoughts and decisions in response to given questions; it must lay out how they interact; it must say why some are important to outcomes in some settings and not in others; and it must be able to tell us how and why things would have been different, given this or that variation in the starting conditions. But given epistemological holism and our actual rationality, what all this amounts to is precisely a general and systematic theory of relevance. (Heal 1996, 83)

The worry for theory theory here is that in the case of human psychology, flexibility and unusual circumstances are the norm.