

## TWO APPROACHES TO MENTAL CONTENT

NENAD MIŠČEVIĆ  
Filozofski fakultet u Zadru  
*Faculty of Philosophy in Zadar*

UDK/UDC: 1  
Izvorni znanstveni rad  
*Original scientific paper*

Primljeno 1990-01-15  
*Received*

Two kinds of approaches to belief contents are contrasted: the davidsonian interpretation theory, and more naturalistic accounts in terms of causal-informational links or proper functioning. It is argued that interpretation theory is a relatively shallow one, aiming at »saving phenomena«, not explaining them, and that it should be revised and grounded in naturalistic account.

Donald Davidson has taught a generation of philosophers how to talk about interpretation. His ideas about using decision-theoretical methods in interpretation, about assumptions concerning rationality and truthfulness of those whose behaviour one interprets, and about psychological and ontological consequences of constraints of interpretation have become a part of common philosophical heritage.

In this paper I would like to compare two approaches to interpretation and ascription of content. One is the classical davidsonian approach, focused on language and meaning and employing the principle of charity as inviolable methodological principle. I shall call this approach Theory of Interpretation, IT for short. The second approach is more naturalistic and evolutionary, focusing on teleologically conceived proper functions of mental states and contents, and I shall accordingly call it Naturalistic Teleological Theory, for short NTT. Although Davidson himself is the initiator of the »classical«, IT approach, he is in his more recent publications stressing the role of more naturalistic constraints (like reliability of perception), and inquiring deeper into causal mechanisms (in his essays on psychoanalysis).

I shall try to defend two claims. The first, which I think is rather uncontroversial is that there are significant parallels between IT and NTT. The second is that NTT, the naturalistic theory, might have some important advantages over its rival, the IT. This second claim is much more controversial.

Let me first list the important similarities:

a) Both theories endorse the same ideal of rational explanations, or explanation by reasons. They both face the same standard problems about the role of reasons in the network of causes, about relation between desire, belief and intention and so on.

b) Both theories endorse a content-oriented views of rational explanation: beliefs and desires explain behaviour largely in virtue of their content.

c) Both theories tend to a moderate version of externalism in defining content-content is defined primarily through »objective«, truth-conditions, not through internal causal roles or sensory input, sense-data and similar purely internal items.

d) Both theories start with fairly optimistic view of agents as being typically well attuned (adapted) to their worlds, and both codify this view in strong principles of optimality – either principle of charity or some principle of teleological wellfunctioning.

e) Both theories share the view of agents as rational. In IT this is a constraint on ascription of mental states, whereas in more functionalistic NTT the rationality-assumption is even stronger: a paralel between logical and functional links is a fundamental postulate.

(I am not well equipped to go into historical exarvations about who influenced whom, so I only note in passing that some of the most important naturalistically oriented authors like Dennett are clearly deeply influenced by Davidson's writings. If it should turn out that there is a systematic influence of IT on theorists of NTT, I would welcome this as an indirect corroboration of the picture I am proposing).

The similarities between IT and NTT are important, because they concern fundamental traits of two theories. There are authors (like D. Papineau in his recent book) which pass over them in silence, stressing only differences. It is therefore useful to state and briefly defend the claim that there are parallels between the two theories.

## DIFFERENCES BETWEEN IT AND NTT

The first and the crucial difference concerns the focus of explanation, or the focus of questions which each theory takes to be relevant.

The typical question asked by IT is the question concerning conditions of intelligibility (for short inteligibility-questions):

What mental properties should be ascribed to the agent if one wants to understand her behaviour?

Although intelligibility-questions are about properties of the agent, they are asked from the perspective of the ascriber in the sense that it is the intelligibility-for-the ascriber which sets constraints on the range of properties ascribed to the agent.

Contrast this with the typical question asked by NTT.

Given that the agent has such-and-such mental properties, how is his having them to be explained? (why-questions)

In the case of why-question, the constraints are set by our knowledge of the make-up of the agent, of possible evolutionary histories and learning histories. By limiting itself to intelligibility-questions the IT opens up the possibility of the following predicament: Suppose we know that in order to understand why the agent did A, we have to ascribe to him a mental set-up M, but we have no idea whatsoever about how in the world the agent came to exemplify M.

So the mathematical theory about regularities of movements of planets, even if approximately correct, needs to be supplemented by, and even integrated into a larger, in this case, physical theory about forces and causal influences.

The IT is good only at systematising the behavioural phenomena, and propogig a unifying framework of coherent mental states giving rise to these phenomena.

Compare this with astronomy: suppose Kepler has just found out that in order to fit the precise measurement data obtained by astronomers, we have to ascribe to the planets movement in elipsoidal orbits with the sun in one focus. Now, even if the theory were true, we would be at loss about explaining the movement: why in the world do planets prefer the less perfect orbit instead of the circle. What we want to know is what causes the planet to take their extravagant routes.

By concentrating on causes, mechanisms and causal histories, the N'TI' offers at least a promise of a deeper explanatory theory.

But, the objector exclaims, is not the IT (together with decision theory) in itself already an explanatory theory!? Does it not explain behaviour by rationalising it? Yes, of course it does. It systematises the data, postulating a hidden order, and allowing us »to view others, nearly enough, as like ourselves« in Davidson's memorable phrase. However, this is only the first step in explanation, not its end. To continue with astronomical analogy, Kepler's theory explains the apparent motion of planets by postulating a simpler but hidden motion. Our curiosity then takes one more turn – why just this kind of motion?

Let us take a closer look at our list of similarities, and note the points where the divergence emerges:

a1) Although both theories accept the davidsonian principle that reasons are causes, they diverge on a deeper level. For the classical IT, reasons just happen to be causes - some pair of events  $E_1$  and  $E_2$  (say the activation of some belief and the onset of an appropriate desire) rationalises a decision  $D$ . This same pair  $E_1$  and  $E_2$  viewed as physical events, will be causes of  $D$ , viewed as physical event. However, for IT the rationalising role of  $E_1$  and  $E_2$  is not systematically linked with their causal role. In contrast, for NTT the two roles are just two aspects of the same process -  $E_1$  and  $E_2$  rationally explain  $D$  in virtue of causing  $D$  in a certain way.

b1) The IT does not bother to ask a very difficult question: how is it possible that content of a mental state should play an explanatory role? Suppose that I want an apple and having appropriate beliefs I reach for one. The IT theorist »saves the phenomenon« by simply assuming that a want for an apple produces reaching for an apple, if circumstances oblige. The assumption is a natural one, and insofar and only insofar an intelligible one (it reminds me of wittgensteinian »logical connection« story). However, unless the theoretician can show at least in principle, how the mental states having content »p« can causally contribute to actions aiming at p, the theory will have to postulate coincidences instead of explaining regularities.

c1) The foregoing applies *mutatis mutandis* to the issue of truth conditions. What makes belief states differentially sensitive to external circumstances? Where does a belief get its content from?

The IT explains why we have to assume that some agent's belief has a content »p«. It does not answer the agent-centered question: in virtue of what does the agent believe that p?

Now, some IT - classicists reply that this agent-centered question is not properly a philosophical one, but a question for neuro-scientist or psychologist. This will not do. Suppose for the sake of argument that the answer is right, and suppose further that the neuro-science finds the right explanation: The agent believes that p because she has a property N. But then, the IT-answer: we have to ascribe to the agent belief that p in order to make her action intelligible, becomes somewhat redundant, because we have to ascribe to her the belief that p anyway, in virtue of her having property N. Suppose, on the other hand, psychology and that neuro-science end in an impasse is that there is no property N such that agent believes that p because of her having N. Then, we have to ascribe to the agent a belief such that her holding that belief can not be traced to any more basic psychological property. Now, how could a philosopher be indifferent to this kind of finding? Would it not undermine the confidence with which

people spontaneously make ascriptions of mental states, and call in question the very point of doing so?

The NTT, in contrast, attempts to account for content in terms of proper functions of mental states, and or in terms of causal links between the states having content »p« and events having both property of causing p-states, and instantiating properties relevant for the truth of »p«.

d1) Why do we have to suppose that the agents are well attuned to the world? Well, first of all, we have to do this in order to start interpreting at all:

»The reason is that we damage the intelligibility of our readings of the utterances of others when our method of reading put others into what we take to be broad error« (The Method of Truth, p. 294.).

Furthermore, since an omniscient interpreter is in the same situation we are,

»...he perforce finds as much agreement as is needed to make sense of his attributions and interpretations, and in this case, of course, what is agreed is by hypothesis true« (ibid).

Now, why are agents in fact well attuned to the world? This can not be so simple because we have to suppose it to be so. Nor does the useful fiction of an omniscient interpreter help here (nor it was intended to help with this question): omniscience is not omnipotence, and although one can conclude from the fact that an omniscient being believes that p to the truth of p, the fact of being believed by the omniscient being does not explain why it is the case that p (Suppose an omniscient but not omnipotent being believes that helium is chemically inert, and you know about the omniscient beings' belief. This still leaves open the question why is helium chemically inert).

If we take a more general view the following difference also emerges:

From the ontological standpoint, it seems that taken in isolation the IT naturally invites an anti-realistic construal (in spite of Davidson's heroic attempts to keep it realistic). The reasons are the following:

First, seeing that the accent is put on conditions of intelligibility for us the ontologist naturally reacts by interpreting these conditions as merely subjective constraints, or at best as projections of our own schemes of understanding onto a neutral behavioural and/or physical reality.

Second, the transcendental status of principles of interpretations opens a gap between realities of science and those uncovered by interpretation. So if one is a scientific realist, and at the same time ready to accept the framework of IT, one will have to attempt to attenuate the ontological status of entities postulated by IT.

In contradistinction to this, the NTT steers away from both the instrumentalist and the transcendentalist variants of anti-realism. One might

even argue that NTT is more faithful to davidsonian rejection of anti-realism than the classical IT.

### THE IT CORRECTED

If our comparison is on the right track, we have established that the classical IT should be integrated into a broader naturalistic framework.

What kind of changes in the picture of interpretation would this produce?

Let us look at some strong constraints imposed by NTT, and working back upon the interpretation.

This is, however, only the beginning of the story. Once you switch from IT to NTT you get stronger constraints which then act back upon the interpretation.

The first type of cases in which this can happen concerns the case where hypotheses about causal mechanisms and causal histories modify the initial assumptions held by interpreter. A good example is given by Davidson himself in his reconstruction of psychoanalysis. In some typical Freudian cases the interpreter will be unable to save the assumption that the agent is rational tout court. What he can do is to postulate the operation of a mechanism which splits the mental of the agent in two parts. Every part is then, taken in itself, behaving rationally, so that the modified version of rationality assumption fits our agent beautifully.

One can conjecture that the change of context will then change the status of the fundamental assumptions, like principle of charity. In the context of IT they are the incorrigible, transcendental principles, whereas in the context of NTT they might become just most general heuristic rules.

The second type of cases concerns the choice between competing rationalisations. In many non-ideal situations (and maybe even in some ideal ones) the behavioural data will, even under the assumptions of charity, underdetermine possible and plausible rationalisations. The classical IT approach has only one kind of recommendation to make: strengthen your assumption of rationality! However, a look at causal mechanisms, even the most simple ones, might suggest which of the competing hypotheses is most likely to be true.

The useful illustration might be the Chomskyan one: if you construct grammars to account for linguistic data with regard only for simplicity, you might end up with a whole lot of equally powerful and equally plausible candidate grammars. But, if you take into consideration, preferably from

the very beginning, the possibilities of implementation, this will drastically reduce the number of allowable grammars.

Finally, a strong explanatory theory will induce the psychologists and philosophers to look for new data, which might not be discovered if one sticks to the less ambitious theory. To return to the case of psychoanalysis: once the psychologist entertains the hypothesis about two rational systems sharing the space within the single skull, he will search for evidence and bring into light a whole lot of otherwise not conspicuous details like: omissions, slips, and parapraxes. He might also consider using materials which would otherwise be considered irrelevant – the way Freud used jokes or dreams.

*Nenad Mišćević: DVA PRISTUPA MENTALNIM SADRŽAJIMA*

S a ž e t a k

Interpretativni (davidsonovski) pristup mentalnim sadržajima uspoređuje se s naturalističkim. Tvrdi se da interpretativni pristup ima status »plitke« teorije čiji je zadatak da »čuva fenomene« komunikacije i razumijevanja. Ova »plitka« teorija podložna je reviziji i bitnom nadopunjavanju u svjetlu naturalističkih, znanstvenih spoznaja o mentalnom.