

## LOGIC AND CONTENT

NENAD MIŠČEVIĆ

Filozofski fakultet u Zadru

UDK: 1

Izlaganje sa znanstvenog skupa

Primljeno: 1988 - 02-05

It has been claimed that possible world semantics is inappropriate for the representation of mental contents, because of its presumed externalist bias. In the article it is shown that possible world provides a tool flexible enough to serve in the discussions about content.

Does one need logic in doing philosophy of mind or philosophy of psychology?

Some time ago it was thought that epistemic logic — logic of belief and knowledge — and logical semantics for propositional attitudes is almost an obligatory format for work in the philosophy of mind, and the possible world semantics was the most common framework used for this sort of research. Recently, the trend has been reversed, and many researches tend simply to bypass the logical issues involved in their methods and results.

It seems that part of the reason for this change of attitude is the feeling that possible world semantics is too rigid to address the interesting issues in the field. Take the central topic of contemporary debate — the issue of the nature of contents of mental states. It is felt that possible world semantics is too objectivistic, that it necessarily represents our propositional contents in an world-oriented or external object-oriented way, thus prejudging the issue in favor of one side in the debate.

I think that this opinion is unjustified. In this short article I shall try to retell some important stories from the relevant philosophical literature, recasting them in very simple terms of epistemic logic. I hope to show that the debate around the notions of »wide« and »narrow« content can be profitably lead in the language of usual modal logic, and I shall propose in a semi-formal way, a class of models which are suitable for this (noble) purpose.

Here is one way to motivate the idea of »narrow content«. Start from the class of verbs the psychological explanation is concerned with: doing, trying, seeing, perceiving, believing, knowing. All of these tell us something about the agent: »Peter saw an aligator« if true, inform us about Peter. Some of those also tell us something about the world (minus the agent): »Peter broke the window« tells us that the window was broken, »Peter knows that it rains« tells us that it rains. Call the last ones »ontologically committed verbs«, and the ones which do not entail anything

about the world »ontologically not-committal verbs«. »To believe« is not committal, »to know« is committal.

Now, some committal verbs, like »to know« have been for a long time submitted to a kind of analysis which treats »John knows that  $p$ « as equivalent (or even synonymous) with a conjunction of »John believes that  $p$ « » $p$  is true« and of some further sentence imposing additional constraints on John beliefs (that they be justified, or tracking the truth or whatever). Some authors (notably H. P. Grice) have analysed perception verbs in the similar way: if John (veridically) sees a tree, then he is in a perceptual state  $P$ , there is a tree, and there is some further connection between the tree and John's being in  $P$ . Some others have offered a similar analysis of action verb (notably D. Armstrong): doing  $r$  means trying to  $r$  plus succeeding. »Being in  $P$ «, and »trying to  $r$ « are non-committal.

One important motive for these analysis has been the quest for generality: if John sees a cherry tree, and Paul hallucinates a tree of the same appearance, they have something in common, which is thought worth capturing by analysis — they are both in the perceptual state  $P$ , only their surroundings are different. Actual methodologies of research into perception give some support to the analysis — it is supposed that a person who has a visual illusion that  $p$  is in the same perceptual state as a person who would veridically perceive that  $p$ , or else we would never be able to learn about normal perception by studying the mechanism of illusion.

Suppose we can isolate the class of committal psychological verbs taking that-clauses as complements. We can represent those verbs as operating on propositions. Let  $O$  be a schematic letter (for »ontologically committed«) for such verbs. Then the commitment can be easily expressed as a constraint on  $O$ :

1.  $T: O_p \rightarrow p$  is a theorem

e. g. »A believes that it rains« does not entail that it rains. The above-mentioned operation  $O_1$  one looks for its non-committal counterpart  $N_1$ : knowing versus believing, doing versus trying, veridical perception versus non-veridical perception etc.

2.  $N_p \rightarrow p$  is not a theorem.

The sought-after non-committal counterpart should have the following property:

3.  $O_1 p \leftrightarrow N_1 p \wedge 1 \wedge p$

Where  $1$  stands for some non-trivial constraint linking  $O_1$  to  $N_1$ .

In plain philosopher's English, the ontologically comittal operator has to be analysed in terms of its non-committal counterpart plus statement that the argument-proposition is true, plus linking conditions: to know that it rains is to believe that it rains plus it being the case that it rains plus there being some suitable connection between the belief and the state of affairs.

Suppose that we can formulate the logics of O and N as both being classical normal logics, i. e. satisfying the following requirements (where O' stands for both O and N):

- a) O containing classical tautologies
- b) O' ( $p \longleftrightarrow q \rightarrow (O' p \longleftrightarrow O' q)$ )
- c) O<sup>t</sup> (where t is tautology)

and closed under substitution and modus ponens.

Then the difference between O and N stemming from the additional axiom (T) for  $O: O_p \rightarrow p$  would be the familiar one matching the contrast between deontic obligation operator and modal necessity operator. Further, adding new axioms to OL and NL we would obtain a set of O-logics and set of N-logics, the first being familiar KT-logics and the second K-logics. It is obvious that K-logics are *more general*, and in our case this generality is essential for psychologists' preference for N-type operators.

On the level of semantics, we have a nice intuitively acceptable contrast.

Given that  $O_p$  entails p, and given the format of possible world frames, the truth of  $O_p$  entails that p is true in the actual world, i. e. that the accessibility relation be reflexive (or, in neighbourhood terms that  $peV(O_p)$  where V stands for valuation set).

The truth of  $N_p$  does not entail the truth of p, so the accessibility relation need not be reflexive.

An even more general approach would allow different arguments for counterpart operators:

$$4. O_i p \longleftrightarrow N_i q \wedge p \wedge 1$$

Now we can state the fundamental assumption presupposed in theories of narrow content:

- 5. FA : Every  $O_i$  has a counterpart  $N_i$ .

It is then usually claimed that only N-type operators are suitable to figure in psychological explanations.

At this point the natural questions to ask is: given a suitable N, are there any restrictions on p?

The methodical solipsist answers affirmatively: Yes, there are important restrictions. At the very least, »p« should contain only terms whose meaning is accessible to the believer, and it should preferably yield a *de dicto* characterisation of believer's thought.

The science-fiction story which is usually told to present the argument is the Twin Earth story. On Twin Earth everything is the same as here, only the water is replaced by a substance XYZ, called on Twin Earth »water« which is phenomenally indistinguishable from water.

Now John, ignorant of chemistry, thinks a thought »Lo, water!«, and Twin John thinks the thought he would express by saying »Lo, water!«.

How should we represent John's belief content?

The proposal which is now fairly accepted in possible world-tradition, and which I will call Standard Proposal (SP) for short wants us to identify the proposition believed with the set of worlds such that if the proposition  $p$  is expressed by a sentence  $s$  then  $p$  is the of worlds at which  $s$  is true (more generally, for any formula  $f$  the set of worlds which satisfy the formula). An exemplary defence of Standard Proposal is to be found in Stalnakers recent book »Inquiry«.

Let us call set of worlds in which  $s$  is true (set of  $s$  — worlds) proposition  $p$  and write  $p$  for it.

— Now, it is easy to show that this picture does not tell much about issues raised by Twin Earth cases. (B. H. Partee has been one of the first to notice it, witness, Partee, 1979.). We may show its inadequacy more vividly if we supplement the Twin Earth story with one more possibility. On the Earth,  $H_2O$  is called »water« and looks like water, on the Twin Earth we have a substance XYZ which looks like water, is called »water«, but is not water. Let us introduce Third Earth, where, due to some electromagnetic forces,  $H_2O$  looks violet, stinks, and is not good for people. There it is called »water«.

Suppose now that we want to use FS to represent John's understanding of the predicate »is water«.

There will be worlds in which John would correctly identify  $H_2O$  as water. But, not all the worlds with  $H_2O$  are such worlds. Worlds which contain Third Earth are the worlds in which John, confronted with a sample of  $H_2O$  (violet, stinking substance) would say: »Oh, no, this definitely is not water!«.

On the other hand, the worlds which contain Twin Earth are worlds in which John would incorrectly identify XYZ as water. So, if we wanted to represent John's understanding of the predicate »is water« we would get a picture quite different from the standard proposal.

Here are four possible reactions to this situation.

The first reaction is the orthodox one: stick to the Standard Proposal to the bitter end, and try to explain away the Twin Earth intuitions. This is Stalnakers way. It preserves the possible world semantics, and sacrifices a substantial piece of psychology.

The second reaction is to drop possible world semantics as inadequate, and to hope for better solution.

The third reaction is to try to enrich the possible world apparatus in order to capture the subtleties of John's referring to water.

The fourth reaction is to build a possible world model in which the reference to the water is absent, and only phenomenally defined predicates are represented.

The first two reactions are extremist. We shall turn our attention to the possibilities inherent in the third one.

Let us suppose that John's thoughts are about water. This is the intuition most people have, and it is worth preserving. On the other hand, the criteria John uses to identify water are successful only in cases which are close enough to normal Earth situations — they are not successful on Twin Earth nor on Third Earth.

If we want to do justice both to John's success and to fallibility of his criteria we shall have to adopt some version of the Principle of Charity. J. Fodor proposes the following one:

»A rough formulation of the Principle of Reasonableness might go: do not be bloody minded in deciding what universe of discourse sentences and beliefs will be evaluated with respect to.« (Fodor, 1982, p. 111).

Fodor's own proposal is to treat sentences about water as one would treat indexical sentences.

We propose a simple implementation of the Principle of Charity.

Let us call a world  $w$  *epistemically hospitable* if it is like our world in the following respect: if something is  $H_2O$  in  $w$  than it looks like water, and so on. The animals that are genetically bears look like our bear. There are no pseudo-cats which look like cats but are in fact robots manipulated by Martians.

John certainly would not be fooled in an epistemically hospitable world. He would recognize water for what it is.

We should evaluate his beliefs in epistemically hospitable worlds only. The upshot will be a kind of compromise: his beliefs will turn out to refer to water, but, the limitation of his criteria will be shown in limitations on the class of acceptable worlds. (In other worlds, John's success will show on the level of our logical theory, his shortcomings will be encoded on the meta-level, in the formulation of truth conditions.)

So, let us start from a standard frame  $F$  from Standard Proposal:  $F = (W, R)$ , where  $W$  is the set of possible worlds  $w$  (indexed in the usual way), and  $R$  is the accessibility relation on  $W$ , intuitively understood as relating our world  $w_0$  to all »belief worlds«.

We next introduce a selection function  $h$  (to remind of »hospitality«) which selects epistemically hospitable worlds. For every formula  $f$  we want to select all epistemically hospitable.

There are two ways of doing it. One is to relativise hospitality to formulas, and to say that a world is hospitable for a formula  $f$  if it does not contain any hidden snares in respect to entities mentioned in the formula. The simpler way is to take the intersection of  $f$ -hospitable world for all  $f$ , in other words, to take worlds which are hospitable for

all formulas. We take this simpler course (if it turns out that in the end it is too restrictive, one can easily switch to *f*-hospitability).

Our selection function  $h: W \rightarrow P(W)$  selects a subset of epistemically hospitable worlds which are *R*-related to the given world  $w$ . Call this subset *H*.

Call the restriction of *R* to *H*-worlds  $R^H$ .

Now we have a new frame. Call it Hospitable Frame,  $F^H$ .

$F^H = (W, R^H)$

We can now evaluate sentences like »John believes that water is wet« in the Hospitable Frame. Sentence of the form  $B_{ap}$  is true in  $F^H$  if it is true in some model in  $F^H$ .  $B_{ap}$  is true in model  $M^H$  from  $F^H$  if it is true at some world  $w_0$  element of  $M^H$ .

$B_{ap}$  is true at  $w_0$  in  $M^H$  under valuation  $v$  iff  $p$  is true at all  $R^H$  related worlds (all  $w_1$  such that  $w_0 R^H w_1$ ).

A formula is valid in *M* iff it is true under all valuation. A formula is valid in  $F^H$  iff it is true in all  $M^H$  elements of  $F^H$ .

In words, restricting our evaluation to hospitable worlds, we represent the fact that John's thoughts are about water, but we also give due place to the consideration of his recognitional abilities — there are worlds in which John would fail to evaluate his own belief correctly.

Logic should be neutral in respect to competing theories, if it is to be of any help in clarifying them. I hope to have indicated that modal logic and possible world semantics are neutral and flexible enough to allow representation of competing claims in philosophy of psychology, and that this can be done using the most elementary means.

### References

- Fodor, J (1982): Cognitive Science and the Twin Earth Problem, Notre Dame Journal of Formal Logic 2/1982.  
 Partee, B.H. (1979): Montague Grammar, Meaning and Mental Reality, in French, Uehling & Wettstein: Contemporary Perspectives in Philosophy of Language, University of Minnesota Press, Minneapolis

### Nenad Mišćević: LOGIKA I SADRŽAJI

#### Sažetak

Kadkad se tvrdi da je semantika mogućih svjetova neprikladna zapriklad teorijā o mentalnim sadržajima, jer je navodno suviše eksternalistička. U članku se nudi jednostavna verzija ove semantike koja je neutralna između internalizma i eksternalizma, te može poslužiti kao zajednički format za raspravu.