The Reasoner, 2012, 6 (1), pp. 5-6, ISSN 1757-0522

Frege's Puzzle from a Model-Based Point of View

Frege's puzzle about propositional attitude reports can be presented in terms of Superman comics. See for example, Thomas McKay, Michael Nelson (2010: Propositional Attitude Reports, *Stanford Encyclopedia of Philosophy*):

At the beginning, Lois Lane does not realize that Clark Kent and Superman are the same person, and she concludes from her observations that *Superman is strong*, but *Clark Kent is not strong*. Thus, it is true that *Lois believes that Superman is strong,* and that *Lois does not believe that Clark Kent is strong*. But since Clark and Superman are the same person, *Clark Kent = Superman* is true as well.

Now, is the rule *F(x) & x=y → F(y)* valid as a general logical principle? If it is, then, by applying it to true sentences, *Lois does not believe that Clark Kent is strong*, and *Clark Kent = Superman*, we should obtain a true sentence: *Lois does not believe that Superman is strong.* However, the sentence *Lois believes that Superman is strong* is true as well, which is a contradiction. Thus, as a general logical principle, *F(x) & x=y → F(y)* seems to be wrong.

This kind of disorder has caused more than a century of controversy. Let's try one more approach to solving the puzzle.

The model-based approach used below can be traced to Marvin Minsky (1965: Matter, Mind and Models, *Proceedings of IFIP Congress 65*, 1: 45-49). In my (2009: Towards Model-Based Model of Cognition, *The Reasoner* 3(6): 5–6) I presented this "robotic ontology" as follows:

“In my head, I have a *world model* (an incomplete one, incoherent, inconsistent, containing all my knowledge, beliefs, etc.). In this model, other persons are believed to have their own world models. Thus, my world model may contain "models of models", for example, a simplified model of your world model.”

But, despite the possible inconsistency of my world model, I don't wish to admit contradictions like Frege's puzzle into it.

How does Frege's puzzle look from this point of view? At the beginning, Lois' world model includes the axiom *Clark Kent ≠ Superman*. Thus, in Lois' world model, her conclusions that *Superman is strong*, but *Clark Kent is not strong* do not contradict each other. But, as a reader of the Superman comics, I know from the very beginning that Clark and Superman are the same person. Hence, in *my* world model, *Clark Kent is strong*, but Lois believes the opposite. At the end of story, Lois is forced to *change* her world model axioms, and Clark becomes strong in her model, too. No puzzle here!

What could have caused the “puzzlification” of the situation?

The statements *Superman is strong* and *Clark Kent is not strong* belong to Lois' initial world model. In this model, *Superman ≠ Clark Kent*. Of course, Lois will not try replacing Superman with Clark Kent in these statements.

The statements *Lois believes that “Superman is strong”, Lois does not believe that “Clark Kent is strong”*, *Lois believes that “Superman ≠ Clark Kent”*, and *Superman = Clark Kent* belong to the world model of the reader, but the parts of the statements in quotes refer to Lois' initial world model. Of course, the reader will not try replacing Superman with Clark Kent in the statement parts referring to Lois' world model.

Thus, one can run into puzzles only by *confusion* of different world models.

A formal model of the situation can be presented as follows. Let's imagine that all sentences we are interested in belong to some *uninterpreted* formal language plus some suitable system of logic. The world model of some person *X* is represented by a set of axioms, which allows to derive all sentences that *X* believes in. Let's denote this axiom set by *WorldModel[X]*. If our logic includes the principle *Ex contradictione sequitur quodlibet*, then we must assume that WorldModel[X] doesn't contain *known* contradictions. The situation of Frege's puzzle is represented as follows:

$\vdash P[Y1]\ \&\ Y1{=}Y2 \rightarrow P[Y2]$;

$WorldModel[X] \vdash P[Y1]\ \&\ Y1{\neq}Y2\ \&\ \neg P[Y2]$.

Of course, no puzzle here!

The triviality of this solution is due to the purely *syntactical* character of the approach. Namely, let's regard world models *not* as “models of the world” with the world itself as their unique “reference”. Let's consider world models simply as the way that people are thinking and talking about the world. When trying to understand their utterances, let's analyse what people *are thinking to be true*, and not what is true “in fact”.

People are comparing and coordinating their world models. But no “independent jury” can be established for comparing of two models M1 and M2, or for comparing of some model M3 with its target system S3 in the world. Speaking strictly, I only can compare things that are contained in my world model: compare *my models* of the models M1 and M2, or compare *my model* of M3 with *my model* of S3.

As demonstrated above, under this approach, at least some of the puzzles disappear...

A similar formulation is attributed to Niels Bohr: "There is no quantum world. There is only an abstract quantum physical description. It is wrong to think the task of physics is to find out how nature *is*. Physics concerns what we can *say* about nature." – quoted after Aage Petersen (1963: The Philosophy of Niels Bohr. *Bulletin of the Atomic Scientists*, XIX(7): 8–14).

Every utterance comes from the world model of the speaker. More generally, every sentence comes from some kind of world model. It may be the world model of a (real or imagined) person, the world model represented in a novel, movie, scientific book, virtual reality, etc. In principle, even smaller informational units (stories, poems, newspaper articles, jokes, mathematical proofs, video clips, dreams, hallucinations, etc.) may introduce their own “partial world models” as small additions to “bigger” world models (regarded as background knowledge). Sometimes, sentences contain references to other world models. Trying to understand such sentences, we should identify, and keep separated, the world models involved.

Karlis Podnieks
Computer Science, University of Latvia