

Frege's Puzzle from Model-Based Point of View

Frege's puzzle about propositional attitude reports (the substitution puzzle) 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 of the story, 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 *Lois does not believe that Clark Kent is strong*. But, since, in fact, Clark and Superman are the same person, *Clark Kent = Superman* is true as well.

Now, is the rule $F(x) \ \& \ x=y \rightarrow F(y)$ 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. Contradiction. Thus, as a general logical principle, $F(x) \ \& \ x=y \rightarrow F(y)$ is wrong?

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

The model-based approach used below can be traced back to [Marvin Minsky](#) (1965: [Matter, Mind and Models](#), *Proceedings of IFIP Congress 65*, 1: 45-49). Minsky applies the notion of model in a way, that is very natural for a computer scientist. In "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, in part – fictional, containing all my knowledge, beliefs, dreams etc.). And I'm acting according to this model. In this model, other persons are believed to have their own world models (in some respects – different from my model). And they are acting according to their models. I may know these models more or less, and in this way I can predict – to some extent – people's behavior. Thus, my world model may contain "models of models" – for example, a simplified model of your world model."

How does look Frege's puzzle from this point of view? At the beginning of the story, Lois' world model includes the axiom *Clark Kent \neq 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 is believing 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 and Clark are different persons, i.e. *Superman \neq Clark Kent*. Of course, Lois will not try replacing Superman by 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 \neq Clark Kent"*, and *Superman = Clark Kent* belong to the world model of the reader, but the statement parts in quotes refer to (the reader's model of) Lois' initial world model. Of course, the reader will not try replacing Superman by Clark Kent in the statement parts referring to Lois' world model.

Thus, one can run into puzzles only by *confusion* of different world models. It seems, some people regard the statement *Superman = Clark Kent* as true in a stronger sense (as true "in fact") than Lois' belief that *Superman \neq Clark Kent*. When such people try injecting their "superior truth" into Lois' world model, they run into Frege's puzzle.

A formal model of the above solution can be presented as follows. Let's imagine that all sentences

we are interested in belong to some common *uninterpreted* formal language plus some suitable system of logic. The world model of some person X is represented by a set of axioms allowing to derive all sentences that X believes in. Let's denote this axiom set by $WorldModel[X]$. Then, the situation of Frege's puzzle is represented as follows:

$\vdash P[Y1] \ \& \ Y1=Y2 \rightarrow P[Y2]$, i.e. this is true in all world models;

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

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 how people are thinking and talking about the world. When trying to understand their utterances, let's analyze what people *are thinking to be true*, and not what is true “in fact”. Then, as demonstrated above, at least some of the puzzles will 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, and sometimes it may contain references to (speaker's models of) other world models. 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, halucinations, 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.

[Karlís Podnieks](#)

Computer Science, University of Latvia

1054 words