

Propositional Attitudes*

Reinhard Muskens

Verbs such as *know*, *believe*, *hope*, *fear*, *regret* and *desire* are commonly taken to express an attitude that one may bear towards a proposition and are therefore called verbs of propositional attitude. Thus in (1) below the agent Cathy is reported to have a certain attitude—namely that of regret—towards the proposition that is the meaning of the embedded sentence.

(1) Cathy regrets that Jim didn't call her

This immediately raises the question what it is exactly that can be the object of an attitude; what kind of thing is the meaning of a sentence? Attitude reports give us a nice context in which to study this question and thus have a central importance for the semantics of natural language. However, the study of propositional attitudes mainly helps us to see what meanings are *not*. Unfortunately it does not give much clue about what meanings are.

1 Can the Object of an Attitude be a Truth Value?

A widely accepted principle in semantics is that if two sentences have different truth values, they cannot have the same meaning (see Cresswell (1982) for a particularly clear statement of the role of this principle). Could it be that meanings just *are* truth values? Of course a theory to this effect would be rather hard to accept, since intuitively there are myriads of meanings while there are only two truth values, truth and falsity.

Attitude reports provide a way to refute such a theory. Let us contrast sentence (1) with sentence (2) below.

*From: R.E. Asher and J.M.Y. Simpson (eds.), *The Encyclopedia of Language and Linguistics*, Vol **6**, pp. 3380–3383, Pergamon, Oxford, 1993. Also in: P. Lamarque (ed.), *Concise Encyclopedia of Philosophy of Language*, pp. 291–294, Pergamon, Oxford, 1997.

(2) Cathy regrets that Joe didn't call her

Now suppose that in fact neither Jim nor Joe called Cathy. If the meanings of the embedded sentences would simply be their truth values then, since they are both true, they would have the same meaning and Cathy would bear the relation of regret to one of them precisely if she would bear that relation to the other. Thus the theory predicts that (2) follows from (1). This is absurd of course, and we may conclude that meanings are not simply truth values.

Note the structure of the argument. We started by supposing that meanings were to be equated with certain entities. From this it was derived that two given sentences had the same meaning and that therefore bearing some attitude towards one implied bearing the same attitude towards the other. This turned out to be absurd and we concluded that meanings were not the things we had supposed them to be. This sort of argument can be brought up against many proposals about the nature of meaning and it can be used to show that there is no real sameness of meaning in natural language, except of course the trivial synonymy that obtains between an expression and itself.

2 The Possible Worlds Account of the Attitudes

Perhaps the standard view about meaning is the view that the meaning of a sentence is the set of all circumstances in which it is true and the most elegant formalisation of this theory is given in the possible worlds semantics that also underlies modal logic. On this account a verb of propositional attitude denotes a relation between an agent and a set of possible worlds, the set of worlds in which the sentential complement of the attitude verb is true. According to this theory, the meaning of (3) can be formalised as formula (4), an expression that denotes the set of all worlds j such that something is a man in j and likes Cathy in j . The theory also provides a way to formalise propositional attitude reports. Sentence (5), for instance, can be rendered as (6), an expression that has for its extension the set of all worlds where the belief relation obtains between Cathy and (4).

(3) Some man likes Cathy

(4) $\lambda j. \exists x (M(x, j) \wedge L(x, c, j))$

(5) Cathy believes that some man likes her

$$(6) \lambda i. Bel(c, \lambda j. \exists x (M(x, j) \wedge L(x, c, j)), i)$$

This last formula gives the non-specific reading, the reading where Cathy believes that at least one man likes her. One virtue of the possible world analysis of the attitudes is that it allows for a second reading of the sentence in question, the reading where (5) states that there is some particular man of whom Cathy believes that he likes her. This reading can be formalised as (7) below. It is called the *de re* reading of (5), while (6) is its *de dicto* reading (see: DE RE-DE DICTO).

$$(7) \lambda i. \exists x (M(x, i) \wedge Bel(c, \lambda j. L(x, c, j), i))$$

There is another phenomenon that is nicely explained by the possible worlds analysis of the propositional attitudes. Consider the invalid argument (8).

$$(8) \frac{\begin{array}{l} \text{Sue knows that the tallest spy is a spy} \\ \text{Tim is the tallest spy} \end{array}}{\text{Sue knows that Tim is a spy}}$$

Sue may know that the tallest spy is a spy without knowing that Tim, who actually is the tallest spy, is a spy. But how can it be that, Tim and the tallest spy being one and the same person, the open sentence *Sue knows that ___ is a spy* is true of one but not of the other? A possible worlds analysis solves the riddle, for even if Tim and the tallest spy are the same in this world, there may be other worlds where this is not so. Therefore Sue may bear an attitude towards the set of all worlds in which the tallest spy has a certain property without having that same attitude towards the set of all worlds in which Tim has that property (see also: COREFERENCE).

However, it is disputed whether the possible worlds theory can handle all such replacements of one noun phrase by a coreferential one in a context of propositional attitude. In particular this may not be so if the two coreferential noun phrases are names. A standard example here chooses *Hesperus* and *Phosphorus*, two names for the planet Venus, as the pair of coreferential names that are not interchangeable. *Hesperus* is the Evening Star, *Phosphorus* the Morning Star and in ancient times it still was unknown that these two heavenly objects were one and the same planet. This means that the argument given as (9) below is not valid.

- (9)
$$\frac{\begin{array}{l} \text{The Ancients knew that Hesperus was Hesperus} \\ \text{Hesperus is Phosphorus} \end{array}}{\text{The Ancients knew that Hesperus was Phosphorus}}$$

The possible worlds account of the fact that (8) is not valid made crucial use of the possibility that although *Tim* and *the tallest spy* may refer to the same person in the actual circumstances they may refer to different persons in other possible worlds. If we want to analyse the Hesperus-Phosphorus paradox along the same lines we must allow for the possibility that although Hesperus and Phosphorus are in fact the same planet, they are only contingently so. This means that we must allow that there might have been situations in which Hesperus and Phosphorus were not the same planet. But some philosophers, especially Kripke (1972, 1979), have argued forcefully against this possibility and hold that the denotations of proper names cannot vary across possible worlds. If this *rigid designator* view of proper names is indeed correct then the sentence ‘Hesperus is Phosphorus’ is true in all possible worlds and thus has the set of all worlds as its semantic value. Since ‘Hesperus is Hesperus’ has the same value, the first premise and the conclusion of (9) cannot but have the same truth value and the argument is predicted to be valid, contrary to standard intuition.

Hintikka (1962) gives an attractive specialisation of the possible worlds approach that should be mentioned here. In general the possible worlds approach to the propositional attitudes is committed to the view that the intension of a verb of propositional attitude is a relation between an agent, a set of possible worlds and a world of evaluation; it is not committed to any particular analysis of this relation. But such a further analysis can have its own merits and Hintikka proposes to define the relation in terms of a simpler one. The attitude of belief, for example, is analysed in terms of the *doxastic alternative* relation, which holds between an agent x , a world j and a world i intuitively if j is compatible with everything that x believes in world i . The report that x believes p is analysed now as the statement that p holds in all of x ’s doxastic alternatives. Writing $B(c, j, i)$ for ‘ j is one of Cathy’s doxastic alternatives in i ’ we can reformalise sentence (5) as formula (10) now.

$$(10) \lambda i. \forall j (B(c, j, i) \rightarrow \exists x (M(x, j) \wedge L(x, c, j)))$$

It should be stressed that there is no incompatibility between this new formalisation and the one given in (6). The relation between (6) and (10)

is one of specialisation. The belief relation is merely redefined in terms of the relation of doxastic alternatives: $Bel(x, p, i)$ now holds if and only if $\forall j(B(x, j, i) \rightarrow p(j))$ does.

3 The problem of Logical Omniscience

If we compare the possible worlds analysis of meaning with the theory that meanings are simply truth values we find that, where the latter account is extremely coarse-grained in the sense that it distinguishes only two meanings, the former approach individuates infinitely more meanings and in general is much more subtle. But even in the case of the possible worlds theory contexts of propositional attitude can help us to see that the individuation of meanings is still not fine-grained enough. The Hesperus-Phosphorus paradox that was mentioned above is one argument that points in this direction, but it is dependent upon the view that names are rigid designators. This view is not universally accepted, but there are other problems with the possible worlds analysis that also suggest that meanings should be discriminated more fine-grainedly. One hard nut is the problem of so-called *logical omniscience*, the false prediction that if somebody knows that φ he also knows that ψ , where ψ is any sentence that is logically equivalent with φ . In order to see that this prediction is indeed false, the following example, adapted from Moore (1989), may be considered.

Suppose some person called Jones wants to enter a building that has three doors, A, B, and C. The distances between any two of these doors are equal. Jones wants to get in as quickly as possible, without making detours and he knows that if A is locked B is not. Now, if our agent tries to open door B first and finds it locked, there might be a moment of hesitation. The reasonable thing for Jones is to walk to A, since if B is locked A is not, but he may need some time to infer this. This contrasts with the case in which he tries A first, since if he cannot open this door he will walk to B without further ado. The point is that one may well fail to realize (momentarily) that a sentence is true, even when one knows the contrapositive to hold. For a moment (11) might be true while (12) is false.

(11) Jones knows that if A is locked B is not locked

(12) Jones knows that if B is locked A is not locked

It follows that the two embedded sentences cannot have the same semantic value, even though they are logically equivalent on the usual account. The possible worlds analysis, on the other hand, predicts that logically equivalent sentences are true in the same possible worlds and thus have the same semantic value. In particular it predicts that the embedded sentences of (11) and (12) have the same meaning and thus that (11) and (12) themselves have the same truth values.

All reasoning takes time. This means that

(13) Jones knows that φ

need not imply

(14) Jones knows that ψ

even if φ and ψ are logically equivalent. If the embedded sentences are syntactically distinct then, since Jones needs time to make the relevant inference, there will always be a moment at which (13) is true but (14) is still false.

This is a problem for the analysis of the attitudes in terms of possible worlds, but worse even, it seems to be a problem for any analysis that is based on some form of standard logic. All ordinary logics allow logical equivalents to be interchanged, but we see here that contexts of propositional attitude do not admit of such replacements. Many researchers therefore have proposed logics that do not support the full interchangeability of logical equivalents. The Bibliography contains a selection of such proposals.

References

- [1] Barwise, J. and Perry J.: 1983, *Situations and Attitudes* MIT Press, Cambridge, Massachusetts.
- [2] Church A 1950 On Carnap's Analysis of Statements of Assertion and Belief. Reprinted in: Linsky L (ed.) 1971 *Reference and Modality*. Oxford University Press, London pp. 168-170
- [3] Cresswell M J 1982 The Autonomy of Semantics In: Peters S and Saarinen E (eds.) *Processes, Beliefs and Questions*. Reidel, Dordrecht pp. 69-96

- [4] Cresswell, M.J.: 1972, Intensional Logics and Logical Truth, *Journal of Philosophical Logic* **1**, 2-15.
- [5] Fagin, R. and Halpern J.Y.: 1988, Belief, Awareness and Limited Reasoning, *Artificial Intelligence* **34**, 39-76.
- [6] Frege, G.: 1892, Über Sinn und Bedeutung, reprinted in Frege, G.: 1962, *Funktion, Begriff, Bedeutung. Fünf Logische Studien*, G. Patzig (ed.), Vandenhoeck, Göttingen.
- [7] Hintikka J 1962 *Knowledge and Belief. An Introduction to the Logic of the two Notions*. Cornell University Press, Ithaca.
- [8] Hintikka, J.: 1975, Impossible Possible Worlds Vindicated, *Journal of Philosophical Logic* **4**, 475-484.
- [9] Konolige K 1985 Belief and Incompleteness In: Hobbs J R and Moore R C (eds.) *Formal Theories of the Commonsense World*. pp. 359-403
- [10] Kripke, S.: 1972, Naming and Necessity, in Davidson, D. and Harman, G. (eds.), *Semantics of Natural Language*, Dordrecht, Reidel, 253-355.
- [11] Kripke, S.: 1979, A Puzzle About Belief, in Margalit, A. (ed.), *Meaning and Use*, Dordrecht, Reidel, 239-283.
- [12] Levesque, H.J.: 1984, A Logic of Implicit and Explicit Belief, *Proceedings AAAI-84*, Austin, Texas, 198-202.
- [13] Mates, B.: 1950, Synonymity, reprinted in Linsky (ed.), *Semantics and the Philosophy of Language*, The University of Illinois Press, Urbana, 1952, 111-136.
- [14] Moore R C 1989 Propositional Attitudes and Russellian Propositions. In: Bartsch R, van Benthem J F A K and van Emde Boas P (eds.), *Semantics and Contextual Expression, Proceedings of the Sixth Amsterdam Colloquium*, Foris, Dordrecht, 147-174.
- [15] Moore R C and Hendrix G C 1982 Computational Models of Belief and the Semantics of Belief Sentences. In: Peters S and Saarinen E (eds.) *Processes, Beliefs and Questions*. Reidel, Dordrecht pp. 107-127

- [16] Muskens R A 1992 Logical Omniscience and Classical Logic. In: Pearce D and Wagner G (eds.) *Logics in AI*, Lecture Notes in Artificial Intelligence **633**, Springer, Berlin pp. 52-64.
- [17] Muskens R A 1995 *Meaning and Partiality*, CSLI, Stanford.
- [18] Quine, W.V.O.: 1966, Quantifiers and Propositional Attitudes, in *The Ways of Paradox*, New York.
- [19] Rantala, V.: 1982, Impossible Worlds Semantics and Logical Omniscience, in I. Niiniluoto and E. Saarinen (eds.), *Intensional Logic: Theory and Applications*, Helsinki.
- [20] Stalnaker, R. *Inquiry*. The MIT Press / Bradford Books, Cambridge Massachusetts.
- [21] Thomason R H 1980 A Model Theory for Propositional Attitudes *Linguistics and Philosophy* **4** pp. 47-70
- [22] Wansing, H.: 1990, A General Possible Worlds Framework for Reasoning about Knowledge and Belief, *Studia Logica* **49**, 523-539.