

# CHAPTER TWO

## PROPOSITIONAL ATTITUDES

*With the basic framework in place, we now proceed to analyze a number of intensional constructions. We start with the basic possible worlds semantics for propositional attitude ascriptions. We talk briefly about the formal properties of accessibility relations.*

### Contents

---

2.1	Hintikka's Idea	17
2.2	Accessibility Relations	20
2.2.1	Reflexivity	20
2.2.2	Transitivity	21
2.2.3	Symmetry	21
2.3	A Note on Shortcomings	22
	Supplemental Readings	22

---

### 2.1 Hintikka's Idea

Expressions like *believe*, *know*, *doubt*, *expect*, *regret*, and so on are usually said to describe PROPOSITIONAL ATTITUDES. The idea is that they express relations between individuals (the attitude holder) and propositions (intensions of sentences).

The simple idea is that *George believes that Henry is a spy* claims that George believes of the proposition that Henry is a spy that it is true. Note that for the attitude ascription to be true it does not have to hold that Henry is actually a spy.

We might want to be inspired by the colloquial phrase “in the world according to George” and say that *George believes that Henry is a spy* is true iff in the world according to George's beliefs, Henry is a spy. We immediately recall from

According to Hintikka [12], the term PROPOSITIONAL ATTITUDE goes back to Bertrand Russell [32].

the previous chapter that we need to fix this idea up by making space for multiple worlds compatible with George's beliefs and by tying the truth-conditions to contingent facts about the evaluation world. That is, what George believes is different in different possible worlds.

The following lexical entry thus offers itself:

$$(29) \quad \llbracket \text{believe} \rrbracket^{w,g} = \lambda p_{\langle s,t \rangle}. \lambda x. \forall w' \text{ compatible with } x\text{'s beliefs in } w: p(w') = 1.$$

What is going on in this semantics? We conceive of George's beliefs as a state of his mind about whose internal structure we will remain agnostic, a matter left to other cognitive scientists. What we require of it is that it embody opinions about what the world he is located in looks like. His beliefs, in other words, if confronted with a particular possible world  $w'$  will determine whether that world may or may not be the world as they think it is. What we are asking of George's mental state is whether anything, any state of affairs, any event, etc. in  $w'$  is in contradiction with something that George believes. If not, then  $w'$  is compatible with George's beliefs. For all George believes,  $w'$  may well be the world where he lives. Many worlds will pass this criterion, just consider as one factor that George is unlikely to have any precise opinions about the number of leaves on the tree in front of my house. George's belief system determines a set of worlds compatible with his beliefs: those worlds that are viable candidates for being the actual world, as far as his belief system is concerned.

Now, George believes a proposition iff that proposition is true in all of the worlds compatible with his beliefs. If there is just one world compatible with his beliefs where the proposition is not true, that means that he considers it possible that the proposition is not true. In such a case, we can't say that he believes the proposition.

Here is the same story in the words of Hintikka [12], the source for this semantics for propositional attitudes:

My basic assumption (slightly simplified) is that an attribution of any propositional attitude to the person in question involves a division of all the possible worlds (...) into two classes: into those possible worlds which are in accordance with the attitude in question and into those which are incompatible with it. The meaning of the division in the case of such attitudes as knowledge, belief, memory, perception, hope, wish, striving, desire, etc. is clear enough. For instance, if what we are speaking of are (say)  $a$ 's memories, then these possible worlds are all the possible worlds compatible with everything he remembers. [...]

How are these informal observations to be incorporated into a more explicit semantical theory? According to what I have said, understanding attributions of the propositional attitude in question

(...) means being able to make a distinction between two kinds of possible worlds, according to whether they are compatible with the relevant attitudes of the person in question. The semantical counterpart to this is of course a function which to a given individual person assigns a set of possible worlds.

However, a minor complication is in order here. Of course, the person in question may himself have different attitudes in the different worlds we are considering. Hence this function in effect becomes a relation which to a given individual and to a given possible world  $\mu$  associates a number of possible worlds which we shall call the ALTERNATIVES to  $\mu$ . The relation will be called the alternativeness relation. (For different propositional attitudes, we have to consider different alternativeness relations.)

EXERCISE 2.1: Let's adopt Hintikka's idea that we can use a function that maps  $x$  and  $w$  into the set of worlds  $w'$  compatible with what  $x$  believes in  $w$ . Call this function  $\mathcal{F}$ . That is,

$$(30) \quad \mathcal{F} = \lambda x. \lambda w. \{w' : w' \text{ is compatible with what } x \text{ believes in } w\}.$$

Using this notation, our lexical entry for *believe* would look as follows:

$$(31) \quad \llbracket \text{believe} \rrbracket^{w,g} = \lambda p_{\langle s,t \rangle}. \lambda x. \mathcal{F}(x)(w) \subseteq p.$$

We are here indulging in the usual sloppiness in treating  $p$  both as a function from worlds to truth-values and as the set characterized by that function.

Here now are two "alternatives" for the semantics of *believe*:

$$(32) \quad \text{ATTEMPT 1 (VERY WRONG)} \\ \llbracket \text{believe} \rrbracket^{w,g} = \lambda p \in D_{\langle s,t \rangle}. [\lambda x \in D. p = \mathcal{F}(x)(w)].$$

$$(33) \quad \text{ATTEMPT 2 (ALSO VERY WRONG)} \\ \llbracket \text{believe} \rrbracket^{w,g} = \lambda p \in D_{\langle s,t \rangle}. [\lambda x \in D. p \cap \mathcal{F}(x)(w) \neq \emptyset].$$

Explain why these do not adequately capture the meaning of *believe*.  $\square$

We can also think of belief states as being represented by a function  $\mathcal{P}$ , which maps an individual and a world into a set of propositions. From there, we could calculate the set of worlds compatible with an individual  $x$ 's beliefs in world  $w$  by retrieving the set of those possible worlds in which all of the propositions in  $\mathcal{P}(x)(w)$  are true:  $\{w' : \forall p \in \mathcal{P}(x)(w) : p(w') = \mathbf{1}\}$ , which in set talk is simply the big intersection of all the propositions in the set:  $\bigcap \mathcal{P}(x)(w)$ . With this notation, our lexical entry would be:

$$(34) \quad \llbracket \text{believe} \rrbracket^{w,g} = \lambda p_{\langle s,t \rangle}. \lambda x. \bigcap \mathcal{P}(x)(w) \subseteq p.$$

EXERCISE 2.2: Imagine that our individual  $x$  forms a new opinion. Imagine that we model this by adding a new proposition  $p$  to the pool of opinions. So,  $\mathcal{P}(x)(w)$  now contains one further element. There are now more opinions. What happens to the set of worlds compatible with  $x$ 's beliefs? Does it get bigger or smaller? Is the new set a subset or superset of the previous set of compatible worlds?  $\square$

## 2.2 Accessibility Relations

Another way of reformulating Hintikka's semantics for propositional attitudes is via the notion of an ACCESSIBILITY RELATION. We talk of a world  $w'$  being accessible from  $w$ . Each attitude can be associated with such an accessibility relation. For example, we can introduce the relation  $w\mathcal{R}_a w'$  which holds iff  $w'$  is compatible with  $a$ 's belief state in  $w$ . We have then yet another equivalent way of specifying the lexical entry for *believe*:

$$(35) \quad \llbracket \text{believe} \rrbracket^{w,g} = \lambda p_{\langle s,t \rangle}. \lambda x. \forall w' : w\mathcal{R}_a w' \rightarrow p(w') = \mathbf{1}.$$

It is profitable to think of different attitudes (belief, knowledge, hope, regret, memory, ...) as corresponding to different accessibility relations. Recall that the linguistic study of determiners benefitted quite a bit from an investigation of the formal properties of the relations between sets of individuals that determiners express. We can do the same thing here and ask about the formal properties of the accessibility relation associated with belief versus the one associated with knowledge, etc. The obvious properties to think about are reflexivity, transitivity, and symmetry.

### 2.2.1 Reflexivity

Recall that a relation is reflexive iff for any object in the domain of the relation we know that the relation holds between that object and itself. Which accessibility relations are reflexive? Take knowledge:

$$(36) \quad w\mathcal{R}_x w' \text{ iff } w' \text{ is compatible with what } x \text{ knows in } w.$$

We are asking whether for any given possible world  $w$ , we know that  $\mathcal{R}_x$  holds between  $w$  and  $w$  itself. It will hold if  $w$  is a world that is compatible with what we know in  $w$ . And clearly that must be so. Take our body of knowledge in  $w$ . The concept of knowledge crucially contains the concept of truth: what we know must be true. So if in  $w$  we know that something is the case then it must be the case in  $w$ . So,  $w$  must be compatible with all we know in  $w$ .  $\mathcal{R}_x$  is reflexive.

Now, if an attitude  $X$  corresponds to a reflexive accessibility relation, then we can conclude from  $\alpha Xs$  *that*  $p$  being true in  $w$  that  $p$  is true in  $w$ . This property of an attitude predicate is often called VERIDICALITY. It is to be distinguished from FACTIVITY, which is a property of attitudes which *presuppose* – rather than (merely) entail – the truth of their complement.

in modal logic notation:  $\Box p \rightarrow p$

If we consider a relation  $\mathcal{R}'_x$  pairing with a world  $w$  those worlds  $w'$  which are compatible with what  $x$  *believes* in  $w$ , we no longer have reflexivity: belief is not a veridical attitude. It is easy to have false beliefs, which means that the actual world is not in fact compatible with one's beliefs, which contradicts reflexivity.

### 2.2.2 Transitivity

[to be written]

### 2.2.3 Symmetry

What would the consequences be if the accessibility relation were symmetric? Symmetry of the accessibility relation  $\mathcal{R}$  implies the validity of the following principle:

$$(37) \quad \text{Brouwer's Axiom:} \\ \forall p \forall w : w \in p \rightarrow \left[ \forall w' [w \mathcal{R} w' \rightarrow \exists w'' [w' \mathcal{R} w'' \& w'' \in p]] \right]$$

in modal logic notation:  $p \rightarrow \Box \Diamond p$

Here's the reasoning: Suppose  $p$  is true in  $w$ . Pick some arbitrary accessible world  $w'$ , i.e.  $w \mathcal{R} w'$ . Since  $\mathcal{R}$  is assumed to be symmetric, we then have  $w' \mathcal{R} w$  as well. By assumption,  $p$  is true in  $w$ , and since  $w$  is accessible from  $w'$ , this means that  $p$  is true in a world accessible from  $w'$ . In other words,  $\exists w'' [w' \mathcal{R} w'' \& w'' \in p]$ . Since  $p$  and  $w$  were arbitrary, and  $w'$  was an arbitrary world accessible from  $w$ , this establishes (37).

To see whether a particular kind of attitude is based on a symmetric accessibility relation, we can ask whether Brouwer's Axiom is intuitively valid with respect to this attitude. If it is not valid, this shows that the accessibility relation can't be symmetric.

In the case of a knowledge-based accessibility relation (epistemic accessibility), one can argue in this way that *symmetry does not hold*:<sup>1</sup>

The symmetry condition would imply that if something is true, then you know that it is compatible with your knowledge (Brouwer's Axiom). This will be violated by any case in which your beliefs are consistent, but mistaken. Suppose that while  $p$  is in fact true, you feel certain that it is false, and so think that you know that it is

<sup>1</sup> Thanks to Bob Stalnaker (pc to Kai von Fintel) for help with the following reasoning.

false. Since you think you know this, it is compatible with your knowledge that you know it. (Since we are assuming you are consistent, you can't both believe that you know it, and know that you do not). So it is compatible with your knowledge that you know that *not* p. Equivalently<sup>2</sup>: you don't know that you don't know that *not* p. Equivalently: you don't know that it's compatible with your knowledge that p. But by Brouwer's Axiom, since p is true, you would have to know that it's compatible with your knowledge that p. So if Brouwer's Axiom held, there would be a contradiction. So Brouwer's Axiom doesn't hold here, which shows that epistemic accessibility is not symmetric.

Game theorists and theoretical computer scientists who traffic in logics of knowledge often assume that the accessibility relation for knowledge is an equivalence relation (reflexive, symmetric, and transitive). But this is appropriate only if one abstracts away from any error, in effect assuming that belief and knowledge coincide.

### 2.3 A Note on Shortcomings

[ . . . to be written . . . stuff about HYPERINTENSIONALITY ]

### Supplemental Readings

We will come back to propositional attitudes and especially the scope of noun phrases with respect to them, including the infamous DE DICTO-DE RE distinction. See Chapters 6 and 7.

Further connections between mathematical properties of accessibility relations and logical properties of various notions of necessity and possibility are studied extensively in modal logic:

HUGHES, George & CRESWELL, Max: 1996. *A New Introduction to Modal Logic*. Routledge London.

GARSON, James: Winter 2003. "Modal Logic." In Edward N. Zalta (Editor) *The Stanford Encyclopedia of Philosophy*. URL <http://plato.stanford.edu/archives/win2003/entries/logic-modal/>, especially section 7 and 8, "Modal Axioms and Conditions on Frames", "Map of the Relationships between Modal Logics".

---

<sup>2</sup> This and the following step rely on the duality of necessity and possibility: q is compatible with your knowledge iff you don't know that *not* q.

A thorough discussion of the possible worlds theory of attitudes can be found in Bob Stalnaker's work:

STALNAKER, Robert: 1984. *Inquiry*. MIT Press.

STALNAKER, Robert: 1999. *Context and Content*. Oxford: Oxford University Press.

Linguistic work on attitudes has often been concerned with various co-occurrence patterns, particularly which moods (indicative or subjunctive or infinitive) occur in the complement and whether negative polarity items are licensed in the complement.

Mood licensing:

PORTNER, Paul: 1997. "The Semantics of Mood, Complementation, and Conversational Force." *Natural Language Semantics*, 5(2): 167–212. doi:10.1023/A:1008280630142.

NPI-Licensing:

KADMON, Nirit & LANDMAN, Fred: 1993. "Any." *Linguistics & Philosophy*, 16: 353–422.

VON FINTEL, Kai: 1999. "NPI Licensing, Strawson Entailment, and Context Dependency." *Journal of Semantics*, 16(2): 97–148. URL [http://www3.oup.co.uk/semant/hdb/Volume\\_16/Issue\\_02/160097.sgm.abs.html](http://www3.oup.co.uk/semant/hdb/Volume_16/Issue_02/160097.sgm.abs.html). Preprint: <http://web.mit.edu/fintel/www/npi.pdf>.

GIANNAKIDOU, Anastasia: 1999. "Affective Dependencies." *Linguistics & Philosophy*, 22(4): 367–421. doi:10.1023/A:1005492130684.

Neg-Raising, cf. ongoing work by Jon Gajewski on his MIT dissertation.

Interesting work has also been done on presupposition projection in attitude contexts, but this can only be appreciated after you have studied theories of presupposition and context change.

HEIM, Irene: 1992. "Presupposition Projection and the Semantics of Attitude Verbs." *Journal of Semantics*, 9: 183–221.

GEURTS, Bart: 1998. "Presuppositions and Anaphors in Attitude Contexts." *Linguistics & Philosophy*, 21(6): 545–601. doi:10.1023/A:1005481821597.

— THIS PAGE INTENTIONALLY LEFT BLANK —