Pragmatics for Propositional Attitudes

Maria Aloni

Background and Motivations

Propositional attitude reports are analyzed in the framework of modal predicate logic (see [4]). Consider the following classical example from Quine which illustrates a difficulty arising for the standard version of this logic:

(1) a. Philip believes that Cicero denounced Catiline.
    b. Philip does *not* believe that Tully denounced Catiline.
    c. Philip believes that x denounced Catiline.

Suppose sentences (1a) and (1b) are true. What is the truth value of (1c) under the assignment that maps x to the individual d which is Cicero *and* Tully? As Quine observes, the ordinary notion of belief seems to require that although (1c) holds when x is specified in one way, namely as Cicero, it may yet fail when the same x is specified in some other way, namely as Tully. Classical modal predicate logic, in which variables are taken to range over plain individuals, fails to account for this ordinary sense of belief.

In a possible world semantics, these 'ways of specifying objects' can be characterized by means of the notion of an (individual) *concept*, i.e. a function from the set of worlds to the set of individuals. Many authors have observed that if we let variables range over concepts rather than plain individuals, we manage to account for Quine's intuition about example (1).[1] An obvious problem arises though if we let variables range over *all* concepts. The following classical example due to Kaplan illustrates why.

Suppose Ralph believes there are spies, but does not believe of anyone that (s)he is a spy. Believing that spies differ in height, Ralph believes that one among them is shortest. Ortcutt happens to be the shortest spy. The *de re* interpretations represented in (2b) and (3b) are deviant in this situation.

(2) a. Ralph believes that someone is a spy.
    b. $\exists x \Box S(x)$

(3) a. Ralph believes that Ortcutt is a spy.
    b. $\exists x(x = o \wedge \Box S(x))$

But, if we let variables range over all concepts – among them the concept *the shortest spy* –, (2b) and (3b) are wrongly predicted to be true.

Not all concepts seem to be suitable to serve as a value for a variable. If we want to solve Kaplan's problem without automatically regenerating Quine's double vision difficulties we must somehow select the set of suitable ones. This is Quine's and Kaplan's diagnoses of these cases. In an influential analysis Kaplan attempts a concrete characterization of the notion of a suitable concept with respect to a specific subject and an object. The following example due to Andrea Bonomi shows that a solution of this kind which attempts to characterize the set of suitable representations as a function of the mental state of the subject disregarding the circumstances of the utterance is condemned to failure.

Suppose that Swann knows that his wife Odette has a lover, but he has no idea who his rival is. He knows that this person is going to meet Odette the following day at the Opera. He decides to kill him there, and he tells his plan to Theo. Odette's

[1] Other non-classical views on trans-world identification, e.g. Lewis's counterpart theory, would also solve this problem (see [1]). These alternative analysis would also be in need of the pragmatic theory I present in this article.

lover is Forcheville, the chief of the army, and Theo is a member of the staff which must protect him. During a meeting of this staff, Theo (who knows all the relevant details) says (4). A murder is avoided.

(4) Swann believes that Forcheville is going to the Opera tomorrow and wants to kill him there.

Consider now sentence (5) uttered by Theo in the same situation.

(5) Swann believes that Forcheville is Odette's lover.

Sentence (5) strikes us as quite out of place. While (4) is acceptable, (5) would be deviant in this situation. On their *de dicto* readings the two sentences are false for obvious reasons. On their *de re* reading they are true only if the concept *Odette's lover* counts as a suitable representation for Swann of Forcheville. So if we follow Kaplan's strategy we are faced with a dilemma: If in order to explain the inadequacy of (5), we rule out the concept 'Odette's lover', we are unable to account for the intuitive acceptability of (4). A natural way out of this *impasse* would be to accept that one and the same representation can be suitable on one occasion and not on another. Many authors (Crimmins & Perry, vFraassen, Stalnaker, vRooy, etc) have observed the crucial role played by *context* in the selection of the set of suitable representations. Not much however has been said about *how* context operates such selection. This is the topic of the present article. First it makes formally precise how contextual information contributes to our evaluation of propositional attitude reports. Second it investigates the question how one arrives to select the intended meaning of these constructions in a given context.

Proposal

I take variables to range over sets of *separated* concepts. Two concepts are separated if their values never coincide. Different sets of concepts can be selected on different occasions. Although variables always range over the same sort of individuals, these may be differently identified. This style of quantification is adopted in modal predicate logic (see [1] for a sound and complete axiomatization of the obtained semantics).

De re attitude reports are analyzed as quantified modal sentences $\exists x_n \Box \phi(x_n)$ which obtain the standard interpretation with the only exception that x_n is taken to range over the set of concepts pragmatically selected for n, rather than over the domain of individuals. In this way their interpretation is made dependent on the conceptualization of the universe of discourse which is contextually operative.

The question arise how addressees arrive to select the intended domain of quantification while interpreting these constructions in a given context. In what follows this question will be addressed in the framework of Optimality Theory.

Optimality theory In O(ptimality) T(heoretic) semantics, interpretation is ruled by a relatively small number of violable principles ranked according to their relative strength (see [5]). These *soft* constraints help us in selecting a set of optimal candidates from a larger set of candidates.[2] The theoretical goal is the formulation of these principles and their ranking in such a way that the actual interpretation of a sentence in a context is the optimal interpretation according to these constraints.

I propose the following constraints as the principles that guide our interpretations of quantified (modal) sentences in a context:

[2] A certain candidate can be optimal even if it violates a constraint provided all alternative candidates lead to more severe constraint violations. A single violation of a higher ranked constraint counts as more severe than many violations of lower ranked constraints.

ANCHOR (vFraassen, Stalnaker, Zeevat, others) Interpretation should be anchored to the context.

CONV MAXIMS (Grice)

BE CONSISTENT (Stalnaker, vdSandt, Zeevat)

BE INFORMATIVE

BE RELEVANT (Horn, vRooy)

***SHIFT** (Williams, Hoop & Hendriks) Do not shift domain of quantification.

***ACC** (vdSandt, Blutner, Zeevat) Do not accommodate.

ANCH, C.MAX > *SHIFT, *ACC

In what follow I will show that by means of these principles, which are not new and find independent motivation in the literature, we can explain the examples of *de re* belief I have discussed in the previous section. I will start by assuming a *one-dimensional* notion of optimality in which the set of candidates are potential meanings[3] of a single syntactic form (see [5]).

Let us start with Quine's question in example (1) *Does Philip believe that* x_n *denounced Catiline?* We have two possibilities, either (i) concept *Cicero* is in n or (ii) concept *Tully*.[4] In the first case, *yes* would be the true answer to the question, in case (ii) *no*. Our principles do not select a unique optimal candidate in this case – after (1a) and (1b) both concepts are equally salient in that situation –, and this explain the never ending puzzling effect of Quine's example. If (1a) alone had been mentioned, or (1b), then our principles would have selected possibility (i) or (ii) respectively, since the alternative interpretation would have violated ANCHOR or *ACC.

Let us now turn to Kaplan's problem of the shortest spy. As above assume that Ralph is unaware that Ortcutt is the shortest spy. Furthermore Ralph does not believe that Ortcutt is fat. Identification by name is prominent. Consider the following sentences:

(6) Ralph believes that Ortcutt is a spy.

(7) Ralph believes that Ortcutt is fat.

These two sentences have two possible meanings in the described situation represented in (a) and (b):

(a) $\Box\phi(ortcutt)$ (*de dicto*) equivalent to
$\lambda x[\Box\phi(x)](ortcutt)$ (*de re* - naming)

(b) $\lambda x[\Box\phi(x)](the\ shortest\ spy)$ (*de re* - description)

Interpretation (a) can be paraphrased as *Ralph would assent to 'Ortcutt is a spy/fat'*, interpretation (b) as *Ralph would assent to 'the shortest spy is a spy/fat'*.

The following diagrams summarize our OT-analysis of these two sentences. I use (*) to indicate that the interpretation violates the corresponding constraint, and !(*) to indicate a crucial violation. Optimal interpretations are those which do not involve any crucial violation.

(6)	**INF,**	**CONS**	***SH/*ACC**
(a)		(*)	
(b)	(*)		!(*)

[3]In this article meanings are identified with subsets of the relevant alternatives (worlds or other) given by the context.

[4]The two concepts cannot be both in n because they are not separated.

(7)	INF, CONS	*SH/*ACC
(a)	!(*)	
(b)		(*)

Interpretation (a) is optimal for sentence (6), because, although it violates one constraint, namely BE CONSISTENT, the alternative candidate leads to more severe constraint violations. Indeed, content (b) violates BE INFORMATIVE – the sentence would be trivialized –, and the weaker principle *SHIFT – our description of the context suggests as active a domain which does *not* contain the problematic concept. Since BE CONSISTENT and BE INFORMATIVE are assumed not to be ranked in any way, the violation of this lower constraint becomes fatal in this case.

Interpretation (b) results however optimal for sentence (7). Our principles wrongly predict that the unnatural concept *the shortest spy* is part of our domain of quantification in this case. Let us have a closer look at this problem.

An intuitive explanation of why content (b) is not assigned to a sentence like (7) in such a situation, is that a speaker, who would have used such a sentence to express such a content, would not have been cooperative. Indeed, content (b) could have been conveyed in a much more efficient way by means of an alternative form, namely (8).

(8) Ralph believes that *the shortest spy* is fat.

It seems that we have an example of the blocking effects discussed by Butler in [2]. The existence of alternative more efficient expressions for content (b) seems to prevent its selection as preferred interpretation for (7).

The existence of alternative more efficient expressions for one and the same content seem to affect what is chosen as preferred interpretation. () is not interpreted as in (b), because a speaker Content (b) is not selected as interpretation for . In the same article Butler observes blocking effects need not be absolute and can be cancelled under special contextual conditions. An example of *de-blocking* triggered by contextual factors seem to be given by the case of Odette's lover. In that situation Theo could have used sentence (9) instead of (10) to say what he wanted to say and he would have been more cooperative. But a use of the former sentence instead of the latter would have been much less effective given Theo's immediate goals.

(9) Swann believes that *Forcheville* is going to the Opera tomorrow.

(10) Swann believes that *Odette's lover* is going to the Opera tomorrow.

Blocking and de-blocking effect cannot be formulated in the one-dimensional analysis we have used so far in which inputs are given by single sentences and no reference is made to alternative forms that the the speaker might have used. In order to account for these cases we need a *bi-dimensional* notion of optimality (see [2]), in which optimal solutions are searched along two dimensions, rather than one: the one of the addressee and the one of the speaker. In the following section I follow [3] and recast bi-dimensional OT interpretation processes using notions from Game Theory. The pragmatics of attitude reports is formalized in terms of *optimization games* in which speaker and addressee – whose preferences are determined by OT preferences in combination with particular goal-directed preferences– coordinate their choice towards optimal form-meaning pairs.

Interpretations as games An *optimization game* is a strategic game $(N, (A_i)_{i\in N}, (\succ_i)_{i\in n})$ involving two players, S(peaker) and H(earer), $N = \{\text{S, H}\}$. S can choose from a set $A_S = \{F1, F2, ...\}$ of alternative syntactic forms. H can choose from a set $A_H = \{C1, C2, ...\}$ of possible semantic *contents*. Optimality theoretic preferences are used in combination with particular goal-directed preferences to define the preference relations of the two players.

In an optimization game, language users are seen as decision makers. Speakers must decide a suitable form for a content to be communicated. Addressees must choose suitable interpretations for a given representation. Since structural *and* pragmatic factors contribute to determine the players' preference relations contextual information determine the output of these games just like syntactic and semantics rules.

In order to make predictions about the outcome of these games we use the game-theoretic counterpart of Blutner's notion of weak optimality, namely bj-optimality. Here is the algorithm proposed in [3] to compute the bj-optimal solutions of a given game.

1. Profiles which point to a Nash equilibrium are blocked.
2. Remove preferences for blocked profiles until you reach a fixed point.
3. The Nash equilibria of the fixed point are the *bj-optimal* solutions in the original game.

As an illustration consider the simple game depicted in the following two diagrams. Preferences are formulated in terms of payoff function in the matrix on the left, and by means of vertical and horizontal arrows in the one on the right.

	C1	*C2*
F1	(4,1)	(4,5)
F2	(3,2)	(1,1)

	C1		*C2*
F1		→	
	↑		↑
F2		←	

This game has one Nash equilibrium, namely the profile (F1, C2). Consequently profiles (F1, C1) and (F2, C2) are blocked. We can then remove the arrow pointing to the former profile. The resulting game has two Nash equilibria, profiles (F1, C2) and (F2, C1). Since there are no more arrows to be removed on the next step, these two profiles are the bj-optimal solutions of our original game.

Let us see now how our examples of the shortest spy and of Odette's lover can be analyzed in this framework.

I propose to characterize the interpretation problem posed by the example of the shortest spy by means of the following optimization game:

	ort		*spy*
'Ortcutt'	*bj*	→	
			↓
'the shortest spy'		→	*bj*

Speaker chooses the row to be played and hearer chooses the column. S can choose between the two forms in (11) and (12).

(11) Ralph believes that *Ortcutt* is fat.

(12) Ralph believes that *the shortest spy* is fat.

Hearer can choose between the two contents represented in (13) and (14).

(13) $ort \mapsto \lambda x[\Box F(x)](ortcutt)$

(14) $spy \mapsto \lambda x[\Box F(x)](the\ shortest\ spy)$

The preference relations of our players are given by the following OT tables.

‘Ortcutt’	**CONS**	***SHIFT/*ACC**
ort	(*)	
spy		(*)

‘the shortest spy’	**CONS**	***SHIFT/*ACC**
ort	(*)	
spy		

Hearer strictly prefers content *spy* over content *ort* given any syntactic input because *ort* crucially violates BE CONSISTENT. Consistent interpretations are preferred over inconsistent interpretations.

Speaker crucially prefers profile (‘the shortest spy’, *spy*) over (‘Ortcutt’, *spy*) because, as it is clear from the OT analysis, while the latter pair violates two constraints, content *spy* can be conveyed by form ‘the shortest spy’ without any constraint violation. Profile (‘the shortest spy’, *spy*) is indeed Nash-optimal in this game. This implies that profile (‘Ortcutt’, *spy*) is crucially blocked, and, therefore, profile (‘Ortcutt’, *ort*) results bj-optimal in this situation. The problematic content *spy* is no longer optimal for sentence (11). We correctly predict that the unnatural concept *the shortest spy* is not taken to be part of our domain of quantification in that situation.

Let us now turn to the case of Odette’s lover which I propose to represent by means of the following game:

	forch		*lover*
‘Forcheville’		→	*bj*
	↑		
‘Odette’s lover’		→	*bj*

S must choose between the following two utterances:

(15) Swann believes that *Forcheville* is going to the Opera tomorrow.

(16) Swann believes that *Odette’s lover* is going to the Opera tomorrow.

H must choose between the following two interpretations:

(17) $forch \mapsto \lambda x[\Box\phi(x)](forcheville)$

(18) $lover \mapsto \lambda x[\Box\phi(x)](odette\text{'}s\ lover)$

The horizontal arrows of Hearer are determined as in the previous case by our OT preferences depicted in the following two tableaus.

‘Forcheville’	**CONS**	***SHIFT/*ACC**
forch	(*)	
lover		(*)

‘Odette’s lover’	**CONS**	***SHIFT/*ACC**
forch	(*)	
lover		

Note however that the vertical arrows are crucially reversed because of the specific goals of Speaker. Although (16) is better than (15) according to our principles, its use in the described situation is much less effective given Theo’s immediate goal of protecting Forcheville’s life. Theo’s intended meaning for sentence (15) results optimal in this situation.

References

[1] Maria Aloni. *Quantification under Conceptual Covers*. PhD thesis, University of Amsterdam, Amsterdam, 2000.

[2] Reinhard Blutner. Some aspects of optimality in natural language interpretation. To appear in *Journal of Semantics*.

[3] Paul Dekker and Robert van Rooy. Bi-directional optimality theory: An application of game theory. To appear in *Journal of Semantics*.

[4] Jaakko Hintikka. Semantics for propositional attitudes. In Davis, Hockney, and Wilson, editors, *Philosophical Logic*. Reidel, Dordrecht, 1969.

[5] Helen de Hoop and Petra Hendriks. Optimality theoretic semantics. *Linguistics and Philosophy*, 24(1), 2001.