Fodor’s puzzle and the metasemantics of attitude reports

Workshop Naming Matters, Stuttgart
Tillmann Pross prosstn@ims.uni-stuttgart.de
Institute for Natural Language Processing, University of Stuttgart
December 2014

1 Overview

• In this paper, I subject to critical scrutiny the semantic dogma that the meaning of an attitude report finds expression in distinct (disambiguated, structurally distinguished) truth-conditions.

• I argue that the exploration of the ‘readings’ resp. distinct truth-conditions of an attitude report reveals properties of the logical form framework in which the distinct truth-conditions are formulated but not properties of the meaning of the attitude report unless the meaning of an attitude comprises the choice for a certain logical form framework.

• Instead, I propose that the meaning of attitude reports is best captured in terms of a communication-theoretic account of semantic interpretation which emphasizes the role of underspecified or partial semantic representations with respect to inference and commitment, thus opening up new perspectives for the semantic, philosophical and psycholinguistic assessment of attitude reports.

2 Introducing Fodor’s puzzle

2.1 The ‘readings’ of “Adrian wants to buy a jacket like Malte’s”

• There is a tacit agreement among semanticists and philosophers of language, that the meaning of the attitude report in (1) finds expression in several distinguished ‘readings’

(1) Adrian wants to buy a jacket like Malte’s.

• The de re and the de dicto ‘reading’ are usually distinguished by considering a minimal pair of contexts for (1)

• The de re ‘reading’ evaluates as true in (2a) but false in (2b) and vice versa for the de dicto ‘reading’.

(2) a. Adrian has decided to buy a certain jacket but has no idea that the jacket he wants to buy is like Malte’s.
   b. Adrian has not decided which jacket he wants to buy but he wants it to be like Malte’s.

• Adrian’s desire can be roughly paraphrased as in (3a) given context (2a) and as in (3b) given context (2b).

(3) a. There is a jacket like Malte’s which Adrian wants to buy.
   b. Adrian wants to buy something that is a jacket like Malte’s.

• The notion of a ‘reading’ is defined in terms of a structural contrast that emerges in the logical form of (1) in first-order intensional predicate logical from placing the phrase a jacket like Malte’s either outside (4a) or inside (4b) the scope of the attitude verb want.
(4)  a. De re 'reading': \((\exists x)(\text{jacket}(x) \& \text{like-Malte's-jacket}(x) \& \text{wants(Adrian, buy(Adrian,x)))}\)

b. De dicto 'reading': \(\text{wants(Adrian, } (\exists x)(\text{jacket}(x) \& \text{like-Malte's-jacket}(x) \& \text{buy(Adrian,x)))} \)

• Now consider the context for (1) in (5) proposed by Fodor [1970]

(5)  Adrian has decided what kind of jacket to buy but has no idea that the kind of jacket he wants is just like Malte’s jacket. (Fodor [1970, cf. 229])

• There is a natural ‘reading’ of (1) given the context (5) that neither the de re (4a) nor the de dicto (4b) ‘reading’ render correctly.

• The de re ‘reading’ (4a) is wrong for this interpretation of (1) because the quantifier (\(\exists x\)) involved in the phrase a jacket like Malte’s entails that there is some particular jacket of which it is true that Adrian wants to buy it.

• The de dicto ‘reading’ (4b) is wrong because it represents the description like Malte’s jacket as part of the content of Adrian’s desire, which in the given context it is not.

• I refer to the ‘reading’ of (1) in context (5) as the Fodorian ‘reading’ of (1).

• What is the logical form of the Fodorian ‘reading’?

• Fodor [1970, cf. 242] showed that under the transitivity of scope relations, not all of the three conditions (6a) - (6c) imposed on the Fodorian ‘reading’ of (1) by the context (5) can be satisfied by a formula of first-order intensional predicate logic at once:

(6)  a. The noun phrase a jacket like Malte’s must be within the scope of the existential quantifier introduced by a jacket if its variable is to be co-referential with the object of Adrian buy . . .

b. The existential quantifier must be within the scope of the verb wants if it is to express the narrow scope ‘reading’.

c. The noun phrase a jacket like Malte’s must be outside the scope of wants if it is to express the ‘reading’ which is transparent for descriptive content.

• The existence of a ‘reading’ of (1) which is distinct from both the de re and the de dicto ‘reading’ but which cannot be defined with the help of the distinction that defines de re and the de dicto ‘reading’ (scope relationships in first-order intensional predicate logic) constitute what I call Fodor’s puzzle.

• But then, how should we approach the Fodorian ‘reading’?

• Let me illustrate the problem arising with Fodor’s puzzle with a short presentation of some of the proposals that have been made in the literature which seek to identify a logical form on the assumption of a Fodorian ‘reading’.

2.2 Fodor’s puzzle in the literature

2.2.1 Romoli and Sudo [2009]

• Let us start with the context in (7)

(7)  Suppose a store sells some jackets that all look like Malte’s and that Adrian does not know anything about Malte. Assume further that Adrian wants one of those jackets and any of them is an option. [Romoli and Sudo, 2009, 427]
(7) gives rise to a Fodorian ‘reading’ of (1) because although (1) is true in (7) it is neither the case that Adrian wants to buy a specific jacket like Malte’s nor does Adrian know that the options which he takes into account for his buy are jackets like Malte’s.

Romoli and Sudo [2009] propose a logical form of the Fodorian ‘reading’ of (1) in context (7) as in (8).

The logical form in (8) analyzes the Fodorian ‘reading’ of (1) as involving a presupposition of a set $X$ of jackets like Malte’s, where Adrian wants to buy one of the members of the set $X$.

(8) $\exists!X : \text{jackets} – \text{like} – \text{malte's}(X)$ and Adrian wants to buy one of $X$ [Romoli and Sudo, 2009, 435]

### 2.2.2 von Fintel and Heim [2011]

- Now consider the logical form (8) against the background of the context in (9).

(9) Suppose a store offers some jackets that all look like Malte’s and that Adrian does not know anything about Malte. Assume that some of the jackets are on sale while others are not and that Adrian is aware of this. Assume further that Adrian wants one of the jackets on sale and any of them is an option.

- There is a ‘reading’ of (1) in (9) which neither the $\text{de re}$ nor the $\text{de dicto}$ ‘reading’ capture.

- In (10), each predicate is annotated with the world in which it is to be evaluated, providing a formalization of the Fodorian ‘reading’ of (1) according to which there are jackets like Malte’s in the actual world $w_0$ which Adrian wants to buy in all of his desire worlds $w'$.

- Note that this approach does not require – unlike the logical form proposed by Romoli and Sudo [2009] – that Adrian wants to buy any actual jacket like Malte’s but restricts the set of actual jackets like Malte’s which Adrian wants to buy to those jackets singled out by Adrian’s desire worlds.

(10) $\lambda w_0 \text{Adrian wants}_{w_0} [\lambda w' [a – \text{jacket} – \text{like} – \text{maltes}_{w_0}].\lambda x_1 [\text{PRO to buy}_{w'x_1}]]$ [von Fintel and Heim, 2011, 102]

### 2.2.3 Schwager [2009]

(11) Suppose Adrian has seen a picture of a certain green Burberry jacket in a catalogue and wants to buy one. Unbeknownst to Adrian, Malte happens to own exactly such a green Burberry jacket. Unbeknownst to Adrian, the type of jacket in the picture which Adrian has seen is sold out and no further jackets of this type have been produced yet: there are no actual jackets like Malte’s.

- Building on a proposal by Cresswell and von Stechow [1982], Kaufmann puts forward a logical form of the Fodorian ‘reading’ of (1) as in (12).

- In (12), $P$ is a structured proposition and $Q'$ a property which is interpreted outside the context of Adrian’s attitude, i.e. $\text{de re}$. $Q'$ is analyzed as the $\text{res}$ of Adrian’s want (in the context (11) manufactured by Burberry) and $P$ is the proposition $\text{buy a jacket with property } Q'$.

(12) $\text{Attitude}_w(x, \langle P, Q' \rangle)$

- Kaufmann’s proposal combines the logical form in (12) with a replacement principle for the property $Q'$ involved in Adrian’s attitude by a reporting property $Q$ (in (1) like Malte’s jacket).
• The Fodorian ‘reading’ of (1) is then analyzed as a requirement on the relation between the reported property \( Q' \) singled out by Adrian’s desire worlds and the reporting property \( Q \) occurring in the report of Adrian’s attitude: “the reported property can be replaced by a different property (the reporting property) as long as the reported property is a subset of the reporting property at all relevant worlds” [Schwager, 2009, p. 409].

• Formally, this proposal for the analysis of the Fodorian ‘reading’ is captured by the set of constraints on property replacement as in (13).

• Applied to the logical form in (12), the replacement principle (13) predicts the Fodorian ‘reading’ of (1) in (11) to be true if the jackets singled out by the property \( Q' \) involved in Adrian’s desire – that of being made by Burberry – are a subset of the set of jackets singled out by the property \( Q \) – that of being like Malte’s jacket – in all relevant worlds.

\[(13) \quad \text{Attitude}_w(x, \langle P, Q \rangle), \text{iff there is a property } Q' \text{ s.t. at the } w\text{-closest worlds } w' \text{ where } Q(w') \neq \emptyset: \]
\[Q'(w') \subseteq Q(w') \]
\[\text{Attitude}_w(x, \lambda w'P_w(Q')) \text{ is true.} \]

[Schwager, 2009, p. 409]

2.3 Facts and Observations

• Obviously, there exist important differences with respect to how the Fodorian ‘reading’ is approached

• The context used for the evaluation of the logical form of (1) are different with respect to the way in which Adrian is acquainted with jackets he wants to buy.

• The way in which logical forms identify the Fodorian ‘reading’ of (1) are different with respect to employed logical form frameworks:
  – Romoli and Sudo [2009]’s proposal states informal paraphrastic truth-conditions supplemented with a theory of presupposition projection
  – von Fintel and Heim [2011] propose a direct interpretation approach supplemented with covert world variables formalized in first-order intensional logic
  – Schwager [2009] uses structured propositions supplemented with a property replacement principle formalized in second-order intensional logic (quantification over properties as functions from individuals to possible worlds)

• Consequently, the type of want-predicate involved in each logical form is quite different:
  – Romoli and Sudo [2009]’s want-predicate is not formally specified
  – von Fintel and Heim [2011]’s want predicate is a one-place predicate of sets of possible worlds
  – Schwager [2009]’s want predicate is a two-place predicate of agents and structured propositions

• The ontological commitments expressed by the logical forms are different:
  – Romoli and Sudo [2009]’s proposal commits to the existence of a non-empty set of jackets like Malte’s
  – von Fintel and Heim [2011]’s proposal commits to the existence of a non-empty set of jackets Adrian wants to buy
  – Schwager [2009]’s proposal commits to the existence of a non-empty set of properties of jackets Adrian wants to buy

• But despite these dramatic differences none of the approaches is wrong in that it doesn’t capture the ‘reading’ of (1) intended by Fodor’s puzzle.

• How can that be?

• In the next section, I argue that the contradicting observations arise from the confusion that is inextricably connected with the notion of a ‘reading’ and the idea that the meaning of an attitude report consists in its disambiguated and structurally distinguished truth-conditions expressed by logical forms from a certain logical form formalism.
3 What is a ‘reading’?

- If a ‘reading’ of an attitude report is identified with a logical form of an attitude report in a certain logical form framework:
  - then the notion of a ‘reading’ is theory-dependent in that sentences may have ‘readings’ in one framework that they don’t have in others, depending on how coarse- or fine-grained logical forms of a certain logical form framework are with respect to attitude reports.
  - Fodor’s puzzle makes this theory-dependence of the notion of a ‘reading’ explicit in that the ‘reading’ intended by Fodor does not hint at any relation with a certain logical form framework in which a logical form of the ‘reading’ could be defined;
  - One part of Fodor’s puzzle is that in that logical form frameworks in which de re and de dicto ‘readings’ are defined in terms of scope relationship, the logical form of the Fodorian ‘reading’ cannot be defined.
  - The other part is that in that logical framework in which the Fodorian ‘reading’ is defined, the Fodorian ‘reading’ is no longer recognizable (because de re and de dicto are no longer available as distinctions).
  - If we fix the de re/de dicto distinction in terms of different structures, i.e. scope relationships in a logical form framework $L_1$, we cannot fix the logical form of the Fodorian ‘reading’ in $L_1$.
  - If we fix the logical form of the Fodorian ‘reading’ in a framework $L_2$ which provides richer structures, we cannot keep fix the de re/de dicto distinction defined in $L_1$ because $L_2$ does not define readings of attitude reports with respect to different scope relationships as $L_1$ does.
  - But the de re/de dicto distinction defined in $L_1$ was used to identify the Fodorian ‘reading’ ex negativo; so if this distinction is not available in $L_2$, the identity conditions of the Fodorian ‘reading’ in $L_2$ are different than they were in $L_1$, in fact the Fodorian ‘reading’ because it is defined relative to $L_1$ does not exist in $L_2$.
  - Consequently, as the choice of $L_1$, $L_2$ or any other logical form framework is not determined by the meaning of an attitude report itself, the notion of a ‘reading’ of an attitude report does not reveal properties of the meaning of an attitude report.
  - The notion of a ‘reading’ reveals properties of a logical form framework, i.e. the structural distinctions among possible logical forms that can be drawn in this framework
  - And if a ‘reading’ is identified with a distinct logical form, then distinct logical forms aren’t related to the meaning of attitude reports either.
  - Put another way, an immediate upshot of the theory-dependence of the notion of a ‘reading’ is that the more precise and fine-grained a semantic theory is, the more ‘readings’, i.e. distinctions in meaning the theory produces and captures. Consequently, further refinement of semantic theory will not pin down the logical form of an attitude report more precisely (“uncertainty principle”).
  - So, to put a long point short: I deny that the notion of a ‘reading’ is of any help in the analysis of the meaning of attitude reports, instead it confuses properties of the meaning of an attitude report with properties of logical form frameworks.
  - This is because Fodor’s puzzle shows that the structure of attitudes is more complex and quite different from the structure of their reports
  - Ultimately, Fodor’s puzzle elucidates that the tacit agreement that the structure of attitudes is the same as that of their report cannot be right.
  - From this point of view, it is not surprising that none of the proposals in the literature assumes that the meaning of attitude reports equals structural distinctions among truth-conditions.
  - Each of the proposals adopts additional non-truth-conditional machinery independent of the assumption of a certain logical form framework, i.e. Romoli and Sudo [2009]’s presupposition projection principle for wide scope indefinites, world variable/situation economy (to avoid overgeneration in von Fintel and Heim [2011]’s approach, see Keshet [2011]) or Schwager [2009]’s replacement principle.
But none of the proposals addresses the question for how a theory of the meaning of attitude reports is like when it is developed independent of the assumption of a certain logical form formalism.

One way to keep the notion of a ‘reading’ useful would be to use a ‘meta’ logical form framework with the help of which the identity conditions of a ‘reading’ forms across frameworks could be determined (as an answer to the question What is a ‘reading’?).

But if ‘readings’ are defined relative to a framework, so is identity of readings and consequently, the notion of a ‘reading’ cannot correspond to a logical form from a certain logical form framework.

In (Pross [2014]), I tried to avoid the problem of theory-dependence by developing a notion of (identity of) ‘reading’ which appeals to the pre-theoretical notion of equaling a ‘reading’ with a paraphrase of a sentence in a certain context.

My idea was to try setting up identity of ‘readings’ via identical predictions of truth/falsity of logical forms assigned to paraphrases in the same contexts, but this requires to prove everyone working on that ‘reading’ wrong in order to determine identity (so, not a good idea, as you can imagine).

But then the problem is to determine which paraphrases count as ‘reading’ and which don’t.

4 What is the meaning of an attitude report?

Consider (14) (from [van Deemter, 1996, 204]).

(14) Watch out! He’s dangerous.

In order for a hearer to grasp the intended meaning of (14) – to infer from (14) that he or she is in danger and should rush into hiding – no disambiguation of the deictic pronoun he is necessary.

Proponents of semantic underspecification (for an overview see van Deemter and Peters [1996]) maintain that the relevant inference from (14) can be executed on an underspecified semantic representation of (14) in which the disambiguation of the deictic pronoun is left to further specification of (14) in context.

With respect to attitude reports, the underspecification-in-context approach has been successfully applied to the de re/de dicto/de se ambiguity by Maier [2009].

He proposed an underspecified logical form for belief reports, which, when applied to an input context gives the right truth-conditions for the belief report based on the relations of acquaintance provided by the context.

But the purpose of an underspecified representation of (14) is more than just to facilitate disambiguation of a compact representation.

What is of primary importance is that the hearer is able to infer the right consequences on the basis of the underspecified representation.

Reyle [1996] puts it this way:

“[I]n almost all of the cases there is not enough information available to identify exactly one ‘reading’. (It is not even clear that the speaker of the sentence had exactly one ‘reading’ in mind.) But nevertheless, we may accept such sentences as true and will, therefore, use the underspecified representations as premises for our arguments. It is thus not enough to say what the underspecified representations look like and how they may be disambiguated. We also must be able to define a suitable consequence relation and to formulate inference rules for them.”

From this point of view, Schwager [2009]’s account of the Fodorian ‘reading’ puts forward an important insight.

Central to her proposal is that the replacement principle is driven by “the sake of reporting an attitude” [Schwager, 2009, 400] and consequently that “we need a proper pragmatic theory to explain when and why speakers choose to rely on the replacement rule” [Schwager, 2009, 411].
Bringing Reyle [1996] and Schwager [2009] together, the purpose of reporting Adrian’s attitude with (1) and not with Adrian wants to buy a Burberry jacket may be that on the basis of the interpretation of (1), the interpreter is able to infer from (1) that she should revise her own plans for buying a jacket like Malte’s if she wants to avoid buying a jacket like Adrian does.

Or, if Adrian doesn’t want to buy a jacket like Malte’s, telling him that the kind of jacket he wants to buy actually is like Malte’s may allow him to infer from (1) that he should revise his desire.

Or, if you want to buy a jacket like Malte’s as a Christmas present for Adrian but get to know that Adrian wants to buy a jacket like Malte’s right now, then you should revise your plans or try to prevent Adrian from putting his plan into practice.

Or, if you report Adrian’s attitude with (1), you commit to have good reason to do so and also to intend that the inferences that can be drawn by the hearer are relevant to her.

This opens up interesting connections with philosophy: spelling out the purposes of attitude reports has been emphasized in the philosophy of action, where the meaning of attitude reports is e.g. defined in terms of their having an impact on future plans of the interpreter of an attitude report (Bratman [1987]), in terms of their rationalizing action (Davidson [1963]) or in terms of their making action understandable (von Wright [1971]).

Furthermore, focusing on inferences instead of distinct truth-conditions in accounting for the meaning of attitude reports has the virtue of pointing to an interesting connection with the debate surrounding inferential semantics in the sense of Brandom [1994].

To a hearer, (1) may not appear ambiguous at all if all she wants (and can get) from it are certain appropriate inferences.

It is interesting to see that Fodor’s puzzle is not the only problematic case of attitude reports for which conclusions similar to mine have been drawn.

van Rooy and Zimmermann [1996] discuss the interpretation of so-called Hob-Nob-Pronouns and conclude that if the intentional identity interpretation of Hob-Nob pronouns “were merely one of several possible readings, it would be hard to explain why one does not think of it if the sentence is uttered out of the blue”. Instead, van Rooy and Zimmermann [1996] propose that there is a default ‘reading’ – the de re ‘reading’ of a Hob-Nob-Pronoun – and that “all other readings only become available if there is reason to rule out this literal ‘reading’. Inspection of […] examples […] shows that they only seem to work fine when accompanied by a longer text setting up the background that eliminates all unwelcome ‘reading’” van Rooy and Zimmermann [1996, 134].

But what is the default ‘reading’ of an attitude report such as (1)? One problem in the identification of default readings is that semantic theory may suggest a different default ‘reading’ than common wisdom, e.g. [Heim, 1992, 211] proposes that “de re construals are ceteris paribus preferred wherever possible” while “common wisdom certainly has it the other way round: de dicto readings are the unmarked choice” [Heim, 1992, 210].

Koralus [2011] discusses yet a different type of ‘readings’ which are hard to reproduce in terms of structurally distinguished logical forms, the so-called referential-attributive ambiguity: Donnellan [1966]:

“The most surprising conclusion of this paper has been that descriptions may not be ambiguous between referential, de re, de dicto, and intermediate interpretations. If this is correct, then sentences including descriptions have one univocal linguistic meaning, which leaves open whether we should give it a referential, de re, de dicto,

Hob-Nob-Pronouns have their name from the example (15) with which [Geach, 1967, 627] introduced the problem.

(15) Hob thinks a witch has blighted Rob’s mare and Nob wonders whether she (the same witch) killed Cob’s sow.

The problem that (15) exemplifies is that in a context where it is not presupposed that witches do exist, neither the de re nor the de dicto analysis of (15) gives the right truth-conditions, so there must be an additional ‘reading’ of (15) of which its exact truth-conditions are subject to debate.

The referential-attributive ambiguity has been observed to arise in the interpretation of sentences involving descriptions such as (16).

(16) Smith’s murderer is insane.

Depending on the context in which (16) is interpreted, (16) can either be paraphrased as Whoever killed Smith is insane (e.g. when looking at the dead body of Smith) or The person who murdered Smith is insane (e.g. when describing the person who is accused of being the murderer of Smith).
or intermediate interpretation. I suggest that this conclusion gives new support to the notion that the correct theory of descriptions requires a representationalist theory of interpretation.”

- The kind of representational framework of meaning Koralus has in mind, and which is suggested by my argument, too, is an account of meaning which explicitly takes into account the role of the interpreter, as e.g. Discourse Representation Theory (DRT, Kamp et al. [2011]) does.

- Also, attitudes and their representations in DRT are first-class citizens of the theory in that attitude representations are motivated and defined independent of the interpretation of linguistic descriptions of attitudes.

- Also, representational theories of meaning offer a straightforward way to incorporate a notion of underspecification.

- Taken together, the argument in this paper can be considered a long and probably not very straightforward motivation to pursue a representational theory of the meaning of attitude reports, a project which has been central to DRT right from the start and is explicitly developed in a number of papers (e.g. Kamp [1985], Asher [1986], Kamp [1988], Asher and Singh [1990], Kamp [1990, 2003, 2005, 2007, 2011], Kamp et al. [2011]).

**The project of a semantic theory of attitude reports**

- Fodor’s puzzle still clearly depicts the methodological limitations of the current state of the art in the analysis of attitudinal semantics

- The development of a semantics of attitudes and their reports which is able to deal with the pervasive ambiguity resp. underspecification and the cognitive relevance of attitudes and their reports is a methodological challenge which requires a perspective on formal semantics that takes into account that language is not a cognitively isolated phenomenon but stands in close relation to other modules of cognition such as sensing, representation and planning.

Some of the questions central to such a theory of attitudinal semantics are:

- Psycholinguistics: How real are the ‘readings’ of an attitude reports? What kind of inferences can be drawn from attitude reports?

- Formal Semantics: From what kind of semantic representation are these inferences supposed to take off? What is the role of intersubjectivity in the interpretation of representations of attitude reports?

- Philosophy: What kind of attitude of whom is a report of an attitude (ascription vs. description vs. representation vs. . . .; hearer vs. speaker vs. attitude bearer)?
References


Hans Kamp. Representing de se thoughts and their reports. Handout for a talk the University of Massachusetts at Amherst, August 2011.


