

## LAYERED BINDING IN *DE SE* REPORTS: ABSTRACT

**1. Synopsis.** Since Heim (1994), it is known that *de se* attitude reports give rise to problematic phenomena for classical Binding Theory (BT). I argue that we should solve the problem by upgrading our technology for variable binding. Variables occurring in attitude reports carry two different indices: one (basic index) tracks identity facts in the actual world, the other (higher-order index) tracks identity facts in the worlds quantified over by the attitude verb. BT only looks at basic indices, while—in consonance with some recent approaches to attitude reports—attitude verbs bind higher-order indices. I show how the picture can be implemented compositionally and survey the prospects of generalizing it to similarly puzzling BT phenomena in Free Indirect Discourse.

**2. The puzzle.** *De se* attitude reports are problematic for classical BT. Focusing on a version of the problem discussed by Sharvit (2011), consider (1):

- (1) McCain convinced Palin to vote for herself.  
(2) [McCain convinced Palin<sub>2</sub> [1 [*PRO*<sub>1</sub> vote for herself<sub>2</sub>]]]

(1) has a reading on which *herself* is read *de re*. On standard assumptions about *de se* reports (see Chierchia (1989), Anand (2006))), the LF for this reading is given by (2): notice that *PRO* and *herself* must carry different indices, lest the latter receive a *de se* interpretation. Yet standard BT predicts that the indexing pattern in (2) is ungrammatical, as it exemplifies a violation of Condition A (on which a reflexive pronoun must be covalued with a c-commanding NP in its local domain). A converse problem obtains with (3), whose standard LF (on the *de re* reading of *her*) is in (4):

- (3) \*McCain convinced Palin to vote for her.  
(4) [McCain convinced Palin<sub>2</sub> [1 [*PRO*<sub>1</sub> vote for her<sub>2</sub>]]]

(3) is ungrammatical, yet standard BT allows the indexing pattern in (4) (since it satisfies Condition B, on which a non-reflexive pronoun must not be covalued with a c-commanding NP in its local domain). The problem generalizes beyond (1)–(4): analogous issues obtain with subject-control verbs and when the controller of *PRO* is a trace.

**3. Layered binding: basics.** Early solutions to the problem are effectively criticized by Sharvit (2011). Sharvit proposes instead to extend the notion of covaluation used in BT to cover cases where a reflexive pronoun denotes the ‘self’ of an attitude holder. This proposal seems empirically adequate, but contains an element of stipulation: it would be striking if BT made *ad hoc* provisions for *de se* reports. This motivates the search for a smoother and more conservative account.

I start from the following diagnosis. The source of the problem is that variable indices in *de se* reports are used to track two kinds of identity facts. On the one hand, they track which actual individuals the attitude is about. On the other, they track who the attitude holder takes those individuals to be. Problems arise when there is a mismatch between facts of the two kinds.

This points towards a semantic solution: we must track separately the two kinds of identity facts. We can do this by switching to a new technique of variable binding, which I call ‘layered binding’. Layered binding allows us to equip variables with multiple sets of indices. The first set tracks actual identity facts, while the second tracks identity facts within attitude worlds. Hence each variable appearing in an attitude report comes with two indices. For example, the LF of (1) is:

(5) [[McCain convinced Palin<sub>1</sub> [ $\lambda^{(a,b)}$ . [ $PRO_1^a$  vote for herself<sub>1</sub><sup>b</sup>]]]]

Indices in subscript position—call them ‘basic indices’—determine covaluation facts used by BT. Indices in superscript position—call them ‘higher-order indices’—determine, roughly, the ‘mode of presentation’ the attitude holder associates to the objects in her attitude state (more details below).

This immediately solves the problem for BT generated by (1)–(4). (1) is ruled in as grammatical on its *de re* reading. By assumption, covaluation facts are determined by basic indices: hence *herself* is covalued with an NP in its local domain, namely *PRO*. (3) is ruled out for analogous reasons.

**4. Implementation.** Following recent accounts (Cumming (2008), Santorio (2012), Ninan (2012)), I start from the idea that attitude verbs work as assignment-shifters and hence bind all variables appearing in their scope. Variables bound in this way range over counterparts of the relevant individuals in the relevant worlds. Here are the resulting truth-conditions for a sample sentence:

[[Jason<sub>1</sub> hopes to kiss her<sub>5</sub>]]<sup>w,g</sup> = true iff for all  $w'$  compatible with Jason’s hopes, the counterpart of  $g(1)$  in  $w'$  kisses in  $w'$  the counterpart of  $g(5)$  in  $w'$ .

The innovation is that attitude verbs manipulate one of two sets of indices, i.e. higher-order indices. Accordingly, the semantics employs two assignments  $g_x$  and  $g_y$ : each of them handles a different set of indices. Only one of them is shifted by attitude verbs:

[[S believes [that p]]]<sup>w,g\_x,g\_y</sup> = true iff for all  $\langle w', g'_y \rangle$  compatible with S’s beliefs, [[p]]<sup>w',g\_x,g'\_y</sup>

The double layer of indices allows variables that are covalued at the basic level to range over different sets of counterparts. Thus the truth-conditions of (1) are (simplifying, and assuming that the higher-order indices on *PRO* and *herself* are, respectively,  $a$  and  $b$ ):

[[ (1) ]]<sup>w,g\_x,g\_y</sup> = true iff for all  $w'$  compatible with what McCain convinced Palin to do, the  $a$ -counterpart of Palin in  $w'$  votes in  $w'$  for the  $b$ -counterpart of Palin in  $w'$ .

The account generalizes naturally to cases involving quantification and to subject-control verbs.

**5. Extras.** Sharvit (2011) points out that an account of BT phenomena in *de se* reports should generalize to similar, but somewhat different, BT phenomena in Free Indirect Discourse (FID). I conclude by arguing that the prospects for this generalization look encouraging. The basic asymmetry in the data is captured on the assumption that FID operators bind basic rather than higher-order indices.

## References

- Anand, Pranav (2006). *De De Se*. Ph.D. thesis, MIT.
- Chierchia, Gennaro (1989). “Anaphora and Attitudes *De Se*.” In R. Bartsch, J. van Benthem, and van Emde Boas (eds.) *Semantics and Contextual Expression*, Dordrecht, Netherlands: Foris.
- Cumming, Sam (2008). “Variabilism.” *Philosophical Review*, 117(4): pp. 525–554.
- Heim, Irene (1994). “Puzzling Reflexive Pronouns in *De Se* Reports.” Handout for Bielefeld conference, March 1994.
- Ninan, Dilip (2012). “Counterfactual Attitudes and Assignment Sensitivity.” Unpublished manuscript, Tufts University.
- Santorio, Paolo (2012). “Reference and Monstrosity.” Forthcoming in *Philosophical Review*.
- Sharvit, Yael (2011). “Covaluation and unexpected BT effects.” *Journal of Semantics*, 28(1): pp. 55.