Covaluation and Unexpected BT Effects

Yael Sharvit
Pre-publication draft

Abstract

It is argued that Binding Theory (BT) must recognize two types of covaluation: the familiar type, which holds between two NPs when they have the same semantic value, and a new type, which holds between two NPs when one of them denotes an attitude holder and the other the ‘self’ of the attitude holder. This is shown to account for the acceptability of ‘de re’ reflexive pronouns and unacceptability of some ‘de re’ non-reflexive pronouns. Alternative theories, which attempt to preserve the standard BT, are also discussed.

1. The problem

The standard Binding Theory (BT), which consists of Conditions A, B and C, constrains the ways NPs may be covalued, where ‘covaluation’ means sameness of semantic value. Based on unexpected BT effects, we suggest that BT must recognize two distinct types of covaluation.

Unexpected BT effects were first observed in Heim (1994a) and recently discussed in Sauerland (2001) and Charlow (2009, 2010a,b). To get a sense of what ‘unexpected BT effects’ refers to, consider first the familiar contrast in (1).

(1) a. McCain convinced Palin to vote for herself.
    b. *McCain convinced Palin to vote for himself.

The standard Condition A requires a reflexive pronoun (in English) to be covalued with a c-commanding NP in its local domain. Condition A accounts for the acceptability of (1a) as follows: assuming the LF in (2) is the LF of (1a) (where LD marks the local domain of herself), herself is indeed covalued with the local c-commanding PRO, which is controlled by Palin.

(2) [McCain convinced Palin [1 [LD PRO$_1$ vote for herself$_1$]]]

By contrast, the unacceptable (1b) doesn’t have a well-formed LF: himself can only be covalued
with *McCain*, which indeed c-commands it, but is also outside the local domain of *himself*, in violation of the standard Condition A.

There are, however, contexts where (1a) can be uttered felicitously, and yet: (i) the well-formed LF in (2) cannot be the LF that is responsible for this; and (ii) moreover, an LF that violates the standard Condition A captures the correct meaning. An example of such a context is this. Sarah Palin, who is running for president, wakes up from a coma and suffers from severe memory loss: she doesn’t remember that she is running for president and perhaps doesn’t even know who she is. McCain visits her in the hospital, and she says to him: “I don’t know who to vote for.” While the two of them look at a picture of her in the newspaper, he says to her: “You must vote for this woman.” Palin, who doesn’t recognize herself in the picture, says: “You are right; I will vote for this woman. She seems reliable.” We can felicitously report this event with (1a).

This Condition A effect is unexpected, because we cannot attribute the acceptability of (1a) to the LF in (2), where *herself* is a ‘de se’ pronoun (covalued with *PRO*, which corresponds to Palin’s ‘self’). The LF in (2) implies that McCain led Palin to the decision “I will vote for myself”; this is not the case in our scenario. On the other hand, the LF in (3) – where *herself* is a ‘de re’ rather than a ‘de se’ pronoun and is covalued with *Palin* – reflects the relevant reading. But it also violates the standard Condition A for the same reason the LF underlying (1b) violates it: the reflexive pronoun is indeed c-commanded by its antecedent, but the antecedent is not in the local domain of the reflexive pronoun.

(3)  \[McCain convinced Palin_2 \{1 [_{LD PRO_1 vote for herself_2}]\}\]

This paper proposes that Condition A is satisfied by more than one type of covaluation: the familiar type, which holds between two NPs when they have the same semantic value (Type-I covaluation), and a new type, which holds between two NPs when one of them corresponds to the ‘self’ of the other (Type-II covaluation). The LF in (3) is thus ruled in.

Section 2 takes a closer look at unexpected BT effects. Section 3 presents the proposed solution – recognizing Type-II covaluation for the purposes of BT. Section 4 explores the consequences of the proposal regarding additional data: (i) De Re Blocking, (ii) De Se Blocking, (iii) unexpected BT effects in Exceptional Case Marking constructions, and (iv) the contrast
between first and third person pronouns in Free Indirect Discourse. Section 5 discusses some alternative solutions, which do not appeal to Type-II covaluation (specifically, the pure ‘de re’ theory, the subject-deletion theory, and the extensional BT theory), and some of the challenges they face.

2. Unexpected BT effects with ‘de re’ reflexive and non-reflexive pronouns

The definition of Type-I covaluation used here is given in (4) (and is based on a similar definition in Heim 2007).

(4) For any root LF $\alpha$ and context $c$ such that $c$ supplies an assignment $g_c$ and $[\alpha]^{f-g_c}$ is defined: any two NPs (of type e), $\beta$ and $\gamma$, that are dominated by $\alpha$ are Type-I covalued in $\alpha$ relative to $c$ if and only if for any $g' \supseteq g_c$ such that $[\beta]^{f-g'}$ and $[\gamma]^{f-g'}$ are defined, $[\beta]^{f-g'} = [\gamma]^{f-g'}$.

The interpretation function $[.]^{f-g}$ is defined in the standard way (for example, as in Heim & Kratzer 1998), which guarantees that each non-terminal node in the syntactic tree is interpreted by the Functional Application rule, the Non-terminal Nodes rule, the Predicate Modification rule or the Predicate Abstraction rule (given in (5)), and that each terminal node that is a trace or a pronoun is interpreted by the Traces/Pronouns rule (given in (6)).

(5) **Predicate Abstraction:** If $\alpha$ is a branching node whose daughters are $\beta$ and $\gamma$, and $\beta$ dominates only an abstraction index $k$, $[\alpha]^{f-g}$ is that function $f$ such that:

1. Dom($f$) = $\{x: [\gamma]^{f-g[k\rightarrow x]}$ is defined$\}$, and
2. for all $x \in$ Dom($f$): $f(x) = [\gamma]^{f-g[k\rightarrow x]}$

(where $g_{[k\rightarrow x]}$ is that function $h$ such that: (a) Dom($h$) = Dom($g$) $\cup \{k\}$; (b) $h(k) = x$; and (c) for all $k' \in$ Dom($h$) such that $k' \neq k$: $h(k') = g(k')$).

(6) If $\alpha$ is a (reflexive or non-reflexive) pronoun or a trace, and $k$ an index, $[\alpha_k]^{f-g}$ is defined only if $k \in$ Dom($g$). Whenever defined, $[\alpha_k]^{f-g} = g(k)$. 
Context c includes (at least) a world (world(c)), a speaker/author (speaker/author(c)) and an assignment (g_c) supplied by c. Ran(g_c) is a subset of the set of semantic objects that are salient for speaker/author(c) in world(c). For simplicity, let us assume that when α is a root note, \([\alpha]^{c-g_c}\) is defined only if g_c is appropriate for α in the following sense: Dom(g_c) contains at most those pronominal indices that are not bound by an abstraction index in α. Thus, when \([7a]^{c-g_c}\) is defined, t_2 and her_2 are Type-I covalued in (7a) relative to c; when \([7b]^{c-g_c}\) is defined (and assuming D_e, the domain of individuals, is a proper superset of \{g_c(1)\}), t_2 and her_1 are not Type-I covalued in (7b), but she_1 and her_1 are (she/her_1 also comes with a gender presupposition: \([she/her_1]^{c-g_c}\) is defined only if g(1) is a female in world(c)).

\[(7)\]

a. \[\text{no woman}\]
\[\begin{array}{c}
\text{S1} \\
\text{S2} \\
\text{S3}
\end{array}\]
\[\begin{array}{c}
t_2 \\
voted for \quad \text{her}_2
\end{array}\]

\[\text{Dom}(g_c) = \emptyset; \text{for any } g' \supseteq g_c \text{ such that } [t_2]^{c-g'} \text{ and } [\text{her}_2]^{c-g'} \text{ are defined, } [t_2]^{c-g'} = [\text{her}_2]^{c-g'} = g'(2).\]

b. \[\text{[she}_1 \text{ saw that } [\text{no woman } [2 [t_2 \text{ voted for her}_1]]]]\]

\[\text{Dom}(g_c) = \{1\}; \text{for any } g' \supseteq g_c, [\text{she}_1]^{c-g'} = [\text{her}_1]^{c-g'} = g_c(1); \text{but there is at least one } g' \supseteq g_c \text{ such that } [t_2]^{c-g'} \text{ and } [\text{her}_1]^{c-g'} \text{ are defined and yet, } [t_2]^{c-g'} \neq [\text{her}_1]^{c-g'}\]

The standard BT says (though see Chomsky 1981 for a slightly different rendition) that a reflexive pronoun must be covalued with some c-commanding NP in its local domain (Condition A); a non-reflexive pronoun cannot be covalued with any c-commanding NP in its local domain (Condition B). ‘Covalued’ means Type-I covalued. Accordingly, (8) and (9) make correct predictions.

\[(8)\]

a. \[\text{[Palin / Every female politician } [6 [\text{LD } t_6 \text{ voted for herself/}^*\text{her}_6]]]\n
---

1 Independent indexing constraints rule out subtrees where one pronominal index has bound and free occurrences.
b. [John was upset because [no woman [Lg t voted for him/*himself]]]]

(9) a. [Lg Palin voted for herself/her]
   b. [John voted for [Lg his mother]]

(9a), with her, is ruled out in any context c where gc(6) = Palin; (9a), with herself, is ruled out in any context c where gc(6) ≠ Palin. However, to simplify matters, we will assume that non-pronominal NPs also carry ‘pronominal’ indices, and that appropriate contextually supplied assignments are one-to-one (but see Heim 2007 for a restriction on this requirement). In what follows we omit those extra indices from the LF. In addition, we assume that economy principles (along the lines of Fox 2000 and others) regulate LF-movement and variable-binding.

Now consider the Object Control McCain convinced every female politician to vote for herself/her, on the reading where herself and her are ‘de se’; i.e., they are co-indexed with PRO. The standard Condition A rules in the LF in (10a), and the standard Condition B rules out the LF in (10b) (well-formed variants of (10a,b) are ones where PRO itself undergoes local QR, cf. (8)); leaving behind a trace).

(10) a. [every politician-w0 [Lg [McCain convinced-w0 t0 [v [Lg PRO5 vote for-w3 herself]]]]]
   b. *[every politician-w0 [Lg [McCain convinced-w0 t0 [v [Lg PRO5 vote for-w3 her]]]]]

(10) is based on the assumption that nominal and verbal predicates take pronominal world-denoting arguments and convince is a propositional attitude verb, which introduces an abstraction index that binds free individual-denoting pronouns (here, PRO5 and herself/her) and an abstraction index that binds free world-denoting pronouns (here, w3). We say, from now on (and following von Stechow 2003), that these pronouns are bound by convince. Convince itself takes an individual-denoting direct object (we call the individual itself ‘object’) and a centered-proposition-denoting (or, perhaps more accurately, a property-denoting) indirect object, the clausal complement.2 Convince quantifies universally over the doxastic alternatives of the

2 Some additional assumptions: (a) when g and gc agree on j, [wj]gc = world(c); (b) verbs that take two objects (e.g., convince) are mapped onto Larsonian trees (see Larson 1988), roughly as in (i).

(i) [Lg [McCain [convincd-w0 [t0 [v [Lg PRO5 vote for-w3 herself]]]]]]
‘object’ – the worlds the ‘object’ doesn’t rule out as candidates for the world she lives in, and the individuals she doesn’t rule out as candidates for who she is – as implied by (11a). The interpretation for (10) is (11b).

\[(11)\]
\[\text{For any } w \in D_s, y \in D_e, x \in D_e \text{ and } p \in D_{\langle\langle\langle,\rangle\rangle}, \llbracket[convince]^e\rrbracket(w)(y)(p)(x) \text{ is defined only if for all } <w', y'> \in \text{Dox}(w, y), p(w')(y') \text{ is defined. Whenever defined, } \llbracket[convince]^e\rrbracket(w)(y)(p)(x) = \text{True iff for all } <w', y'> \in \text{Dox}(w, y), p(w')(y') = \text{True.}\]

\[\text{b. Whenever defined, } \llbracket(10a,b)\rrbracket^e = \text{True iff for every } x \in D_e \text{ such that } x \text{ is a (female) politician in world(c), for all } <w, y> \in \text{Dox(world(c), x)}, y \text{ votes for } y \text{ in } w.\]

The standard BT indeed makes correct predictions regarding (10), as the contrast in (12) shows.

\[(12)\]
\[\text{Context: Yesterday, McCain visited three different female politicians – Sarah Palin, Hilary Clinton, and Caroline Kennedy. Knowing that they all considered self-voting unethical, he talked to them about changing their minds.}\]
\[\text{a. McCain must have convinced every female politician to vote for herself, because each of them is now mumbling to herself: “I will certainly vote for myself.”}\]
\[\text{b. *McCain must have convinced every female politician to vote for her, because each of them is now mumbling to herself: “I will certainly vote for myself.”}\]

**Interpretation:** For all female politician x, McCain convinced x to self-vote.

---

The extra verbal layer is often omitted in the text, but the readers should bear in mind that the NP object of *convince* is assumed to e-command the clausal complement of *convince*.

This is obviously a very simplified semantics. The actual semantics of the Object Control *convince* probably implies that the ‘subject’ said something or did something that led the ‘object’ to conclude/decide the content of the complement clause (cf. Comrie 1984; Sag & Pollard 1991; Stephenson, to appear). For our purposes the simplified semantics in (11a) suffices.

We assume (with Schlenker 1999, von Stechow 2003 and many others), that when *herself* is bound by *convince*, its phi-features may be invisible to semantic interpretation whenever the controller has the same phi-features, as *McCain convinced Palin to vote for herself* is felicitous even when Palin is unaware of the fact that she is a female. The ungrammaticality of *McCain convinced Palin to vote for himself*, reporting a ‘de se’ thought when Palin takes herself to be a man, is explained by the assumption that the phi-features on *himself* (including 3rd person) must be visible to semantic interpretation, because *Palin and himself* do not agree in phi-features.
Now, what predictions does the standard BT make regarding (13), where *herself* and *her* are ‘de re’, rather than ‘de se’, pronouns (i.e., they are not bound by *convinced*)? (13a) is predicted to be ruled out because *herself* is free in its local domain, in violation of Condition A. On the other hand, (13b) is predicted to be ruled in because *her* is free in its local domain, satisfying Condition B.

(13) a. *[every politician-*w* ] [McCain convinced-*w* t* ] [LD PRO* vote for-*w* herself*]]

   b.   [every politician-*w* ] [McCain convinced-*w* t* ] [LD PRO* vote for-*w* her]]

Let us examine what (13a,b) are supposed to mean, assuming an appropriate theory of ‘de re’ expressions (e.g., Lewis 1979, von Stechow 1984, Heim 1992). According to one particular version of this theory (e.g., Abusch 1977, Cresswell & von Stechow 1982), the LFs in (14) are actually the LFs that underlie the ‘de re’ readings.

(14)a. *[every politician-*w* ] [McCain convinced*de-re-*w* t* ] [LD PRO* vote for-*w* herself*]]

   b.   [every politician-*w* ] [McCain convinced*de-re-*w* t* ] [LD PRO* vote for-*w* her]]

It is assumed that [[*convince*]c,g] in (14) takes as its third internal argument, not the centered proposition denoted by its clausal complement, but rather the corresponding structured centered proposition – a pair whose first member is the individual denoted by *herself/her*, and whose second member is a member of D<e,<s,<e,t>> – as shown in (15a). The semantics of *convince* is given in (16), which implies that the context supplies a salient suitable (acquaintance-based) function from world-individual pairs to individuals (we call it, informally, a salient suitable description), which picks out the ‘res’ – in our case, the denotation of *herself/her* (the individual whom the attitude is about) – in the actual world, and may pick out some other individual in the doxastic alternatives of the attitude holder. (15b,c) provide the interpretation of this LF.

---

5 For alternative analyses of ‘de re’ attitudes, which make the syntax-semantics mapping more transparent, see Heim’s (1994b) ‘res’-movement analysis, Percus & Sauerland’s (2003a) concept-generator analysis, and others.

6 To be precise, the “name” of the contextually-supplied description in (16) should be F<z,F> (but we simplify it here). Notice that not all variants of this approach rely on the assumption that the context supplies a unique description (see, for example, Cresswell & von Stechow 1982, Percus & Sauerland 2003). A possible alternative semantics for *convince* is this. [[*convince*]c,g](w)(z)(<y, f>)(x) is defined only if: (i) there is at least one F ∈ C<z> such that
(15) a. For any context $c$ such that $\llbracket (14a,b) \rrbracket^c_D$ is defined and any $g' \supseteq g_c$ such that $6 \in \text{Dom}(g')$, the structured centered proposition corresponding to the clausal complement of $\text{convince}^{\text{de-re}}$ is: $\llbracket g'(6), \llbracket \llbracket 7 \ 3 \ 5 \ [\text{PRO}_3 \ \text{vote-for-w}_3 \ t_3] \rrbracket^c_D \rrbracket \rrbracket^c_D$.

b. For any context $c$ such that $\llbracket (14a,b) \rrbracket^c_D$ is defined (at the very least, for every $x \in D_e$ such that $x$ is a (female) politician in $\text{world}(c)$, $F_c^x(\llbracket t_5 \rrbracket^c_D(6 \rightarrow x)) = \llbracket \text{herself/her}_6 \rrbracket^c_D$), the structured centered proposition corresponding to the clausal complement of $\text{convince}^{\text{de-re}}$ is: $\llbracket \llbracket 14a,b \rrbracket^c_D \rrbracket^c_D = \text{True}$ iff for every $x$ such that $x$ is a (female) politician in $\text{world}(c)$, for all $<w, z> \in \text{Dox}(\text{world}(c), x)$, $z$ votes for $F_c^x(z)(w)$ in $w$.

c. An example of a possible value for $F_c^x$ (for some $x$): $[\lambda z \in D_e . \lambda w \in D_s : \text{there is exactly one } y \in D_e \text{ such that } z \text{ is looking at } y \text{ in } w . \text{ the unique } y \in D_e \text{ such that } z \text{ is looking at } y \text{ in } w]$

(16) For any pair $<y, f>$ (where $f \in D_{<e, <s, e, d_s>>}$ and $y \in D_s$), $z \in D_e$, $x \in D_e$, and any $w \in D_s$, $\llbracket \text{convince}^{\text{de-re}} \rrbracket^c_D(w)(z)(<y, f>)(x)$ is defined only if $c$ furnishes a $<e, <s, e, d_s>>$-function, $F_c^y$, such that $F_c^y \in C_z$, $F_c^y(z)(w) = y$ and for all $<w', z'> \in \text{Dox}(w, z)$, $f(F_c^y(z')(w'))(w')(z')$ is defined (where $C_z$ is the set of suitable $<e, <s, e, d_s>>$-functions relevant for $z$ in $c$). Whenever defined, $\llbracket \text{convince}^{\text{de-re}} \rrbracket^c_D(w)(z)(<y, f>)(x) = \text{True}$ iff for all $<w', z'> \in \text{Dox}(w, z)$: $f(F_c^y(z')(w'))(w')(z') = \text{True}$.

It is further assumed that the context supplies as many descriptions as there are ‘res’-denoting expressions (and that the argument of $\text{convince}^{\text{de-re}}$ that corresponds to its clausal complement adjusts accordingly).

The predictions of the standard BT – namely, that (13a)/(14a) violate Condition A whereas (13b)/(14b) respect Condition B – are wrong, as evidenced by the contrast in (17). In accordance with what is observed in Heim (1994a), and with speakers’ intuitions regarding (17a), (14a) should be ruled in. We add the observation that according to speakers’ intuitions concerning

\[F(z)(w) = y; \text{ and (ii) for all } F \in C_z \text{ such that } F(z)(w) = y, \text{ for all } <w', z'> \in \text{Dox}(w, z), f(F(z')(w'))(w')(z') \text{ is defined (where } C_z \text{ is the set of suitable } <e, <s, e, d_s>>-\text{functions relevant for } z \text{ in } c). \text{ Whenever defined, } \llbracket \text{convince}^{\text{de-re}} \rrbracket^c_D(w)(z)(<y, f>)(x) = \text{True} \text{ iff there is at least one } F \in C_z \text{ such that } F(z)(w) = y \text{ and for all } <w', z'> \in \text{Dox}(w, z): f(F(z')(w'))(w')(z') = \text{True}.\]
(17b), (14b) should be ruled out.

(17) Context: McCain visits three different female politicians – Sarah Palin, Hilary Clinton, and Caroline Kennedy. He knows that they suffer from severe memory loss. He wants to guarantee that Palin’s ballot says “Sarah Palin”, Clinton’s ballot says “Hilary Clinton”, and Kennedy’s ballot says “Caroline Kennedy”, so he gives each one of them a picture of herself. None of them recognizes herself in the picture. After he leaves, we ask each of them who she will vote for. Each one of them, while pointing at her own picture, says: “For no one; certainly not for this woman.” So we know he failed with all of them.

a. McCain didn’t convince any amnesiac female politician to vote for herself, because by the time he gave each one of them a picture of herself and said: “You should vote for this woman,” it was already too late to influence any of them.

b. *McCain didn’t convince any amnesiac female politician to vote for her, because by the time he gave each one of them a picture of herself and said: “You should vote for this woman,” it was already too late to influence any of them.

Interpretation: there is no amnesiac female politician x such that McCain convinced x to vote for x.

As is well known, Condition B may often be obviated in special circumstances. For example, Everyone hates John; even he (himself) hates him, where the subject of the second clause is focused, is acceptable. Another case where Condition B may be obviated is when the preceding discourse makes it possible to interpret the pronoun as a ‘paycheck’ (or E-type) pronoun, as in (18) – an improved version of (17b) (for discussion of the circumstances that allow obviation of Condition B, see Lakoff 1972; Evans 1980; Reinhart 1983; Grodzinsky & Reinhart 1993; Grodzinsky 2007; Heim 1993, 2007; Schlenker 2005; among many others).

(18) McCain showed every amnesiac female politician a picture of a woman, namely, herself. Still, he didn’t convince any one of them to vote for her.

Crucially, (12) and (17) lack the kind of special context required to obviate Condition B.
The unexpected BT effects in (17) are not unique to Object Control verbs such as *convince*, that quantify over the doxastic alternatives of the ‘object’; they arise also with Subject Control verbs such as *promise*, that quantify over the doxastic alternatives of the ‘subject’. The reading in (19a) is available with *himself* but not with *him* (as predicted by the standard BT), and the reading in (19b) is also available with *himself* but not with *him* (contrary to the predictions of the standard BT).

(19) Every male politician promised Palin to vote for himself/*him.

a. The promise: “I will vote for myself”.

*Every male politician promised*-w₀ Palin [3 5 [LD PRO₅ vote for-w₃ himself/*him₃]]

b. The promise: “I will vote for this guy.”

*Every male politician [6 [t₆ promisedᵣₑ-rₑ-w₀ Palin [3 5 [LD PRO₅ vote for-w₃ himself/*him₆]]]]*

Is there is a simple and straightforward explanation for these effects? According to Heim’s (1994b) ‘res’-movement theory, a ‘de re’ expression moves out of the embedded clause and becomes a syntactic argument of the embedding verb. Accordingly, the LFs of *McCain convinced Palin to vote for herself/*her* are as in (20): the standard Condition A is met in (20a) and the standard Condition B is violated in (20b). These are indeed correct predictions.

(20) a. *[Palin [6 [LD McCain convincedᵣₑ-rₑ-w₀ t₆ herself₆ [7 3 5 [PRO₅ vote for-w₃ t₇]]]]]

b. *[Palin [6 [LD McCain convincedᵣₑ-rₑ-w₀ t₆ her₆ [7 3 5 [PRO₅ vote for-w₃ t₇]]]]]

But this explanation should probably be rejected, in view of the fact that it makes wrong predictions for *McCain convinced Palin to vote for him/*himself*. Again, the standard Condition A is met in (20a) and the standard Condition B is violated in (20b), but this time, the prediction is wrong.⁷

---

⁷ One could try to take advantage of the Larsonian structure (see Footnote 2), and claim that in (21) *herself/her* are not “close enough” to the subject. But subjects in double-object constructions do license reflexives and ban covalued non-reflexives (e.g., *John introduced himself/*him to Mary and John introduced Mary to himself/*him*).
There are more promising attempts to explain unexpected BT effects which, like the one we have just considered, do not involve changing basic standard assumptions about how BT operates. Some of these attempts are explored in Section 5. In section 3 we present the current proposal, which relies on a new notion of covaluation, and in Section 4 we explore further consequences and predictions of the proposal presented in Section 3.

3. The proposal

We propose that the Binding conditions appeal to two types of covaluation: Type-I covaluation (the familiar type, defined in Section 2) and Type-II covaluation. Informally, Type-II covaluation holds between two NPs if one of them denotes an attitude holder and the other corresponds to the ‘self’ of the attitude holder. Condition A still says that a reflexive pronoun must be covalued with some c-commanding NP in its local domain; and Condition B still says that a non-reflexive pronoun may not be covalued with any c-commanding NP in its local domain. But ‘covalued’ is now understood as ‘Type-I or Type-II covalued’ (and, as pointed out to me by Alan Prince, if it turns out that yet additional covaluation types are relevant, ‘covalued’ will cover them as well).

Let us first illustrate informally how the proposal works starting with our key example, namely, the Object Control McCain convinced Palin to vote for herself/*her, where herself/*her is ‘de re’. The idea is that PRO and herself/*her are Type-II covalued, in satisfaction of Condition A and in violation of Condition B. PRO and herself/*her are Type-II covalued in (22) because herself/*her₆ are coindexed (and therefore Type-I covalued) with t₆ (which denotes an attitude holder) and PRO₃ – being bound by convince^{de-re}, and being controlled by t₆ – corresponds to the ‘self’ of the denotation of herself/*her₆.

(22) [Palin [McCain convinced^{de-re} w₀ [McCain [McCain convinced^{de-re} w₀ t₆ [3 5 [PRO₃ vote for -w₃ t₆]]]]]]

Likewise, in the LFs underlying the Subject Control McCain promised Palin to vote for himself/him, where himself/him is ‘de re’, PRO and himself/him are Type-II covalued.

On the other hand, McCain convinced Palin to vote for himself is ruled out, and McCain
convinced Palin to vote for him is ruled in. In both cases, since the controller is the trace of Palin, and not McCain, Type-II covaluation does not hold between PRO and himself/him. Finally, McCain expects Palin to vote for himself is ruled out, and McCain expects Palin to vote for him is ruled in. This variety of expect is not a Control verb, and the subject of the embedded clause is Palin, which is not covalued – in any sense – with him/him.

We now introduce a more formal rendition of the proposal. As a first step, let us note that attitude verbs such as believe, convince and promise carry the presupposition that their clausal complement is defined in the doxastic alternatives of the attitude holder. This is illustrated by the semantics of Object Control convince in (11a) and its ‘de re’ variant in (16), which is repeated in (23).

\[(23)\] For any pair \(<y, f>\) (where \(f \in D_{\langle e, s, e, d, e>\rangle}\) and \(y \in D_e\)), \(x \in D_e\) and \(z \in D_e\), and any \(w \in D_s\), 
\[
[[\text{convince}^{\text{de-re}}]]^f(z)(<y, f>)(x) \text{ is defined only if } c \text{ furnishes a } <e, s, e> - \text{function, } F^y_c, \text{ such that } F^y_c \in C_s, F^y_c(z)(w) = y \text{ and for all } <w', z'> \in \text{Dox}(w, z), f(F^y_c(z')(w'))(w')(z') \text{ is defined (where } C_s \text{ is the set of suitable } <e, s, e> - \text{functions relevant for } z \text{ in } c). \text{ Whenever defined, } [[\text{convince}^{\text{de-re}}]]^f(z)(<y, f>)(x) = \text{True} \text{ iff for all } <w', z'> \in \text{Dox}(w, z): f(F^y_c(z')(w'))(w')(z') = \text{True}.\]

It follows that the set of contexts where (22) is defined is a subset of the set of contexts where Palin believes the presuppositions of the clausal complement. Coupled with a slightly more formal definition of Type-II covaluation (see (24)), this results in PRO and herself/her being Type-II covalued in (22).

\[(24)\] For any root node \(\alpha\) and any attitude verb ATT such that \(\alpha\) dominates ATT, let: (i) \(w(\text{ATT})\) be the pronominal world-argument of ATT, (ii) \(S(\text{ATT})\) be the first node that dominates ATT and all its syntactic arguments, (iii) \(R(\text{ATT})\) be the highest node in the clausal complement of ATT that is c-commanded by both abstraction indices introduced by ATT, and (iv) \(k(\text{ATT})\) be the world-binding abstraction index introduced by ATT and \(m(\text{ATT})\) the individual-binding abstraction index introduced by ATT. Then –

Any two NPs (of type e), \(\beta\) and \(\gamma\), that are dominated by root \(\alpha\) are Type-II covalued in \(\alpha\) if and only if there is at least one ordered pair \(<\text{NP1}, \text{NP2}> \in \{<\beta, \gamma>, <\gamma, \beta>\}\) such that:
for any context $c'$ such that $g_{c'}$ is appropriate for $S(ATT)$ and $[[S(ATT)]]^{c'-c}$ is defined, any $g' \supseteq g_{c'}$ such that $\text{Dom}(g') = \text{Dom}(g_{c'}) \cup \{j: j$ is a pronoun that is free in

NP1$\}$ and $[[NP1]]^{c'-c}$ is defined, and any $<w, x> \in \text{Dox}([[w(ATT)]]^{c'-c} , [[NP1]]^{c'-c})$

$$\{g'' \supseteq g': [[7 \quad [R'(ATT)]] \quad ]^{c'-c} \quad (F(x)(w)) \quad \text{is defined}\} \cap \{g'' \supseteq g'_{[k(ATT)-\omega]}: [[NP2]]^{c'-c} = x\} \neq \emptyset$$

(or, given a node $\chi$ dominated by $R(ATT)$ such that $[[\chi]]^{c'-c}$ is a ‘res’,

there is at least one $<e,<s,e>>$-function $F \in C'_{NP1}^{c'-c}$ (the set of relevant suitable

$<s,<s,e>>$-functions in $c'$) such that:

(a) $F([[NP1]]^{c'-c})([[w(ATT)]]^{c'-c}) = [[\chi]]^{c'-c}$, and

(b) for all $<w, x> \in \text{Dox}([[w(ATT)]]^{c'-c} , [[NP1]]^{c'-c})$:

$$\{g'' \supseteq g': [[7 \quad [R'(ATT)]] \quad ]^{c'-c} \quad (F(x)(w)) \quad \text{is defined}\} \cap \{g'' \supseteq g'_{[k(ATT)-\omega]}: [[NP2]]^{c'-c} = x\} \neq \emptyset$$

where:

(i) $R'(ATT)$ is just like $R(ATT)$ except that $t_7$ replaces $\chi$; and

(ii) $<[[\chi]]^{c'-c}, [[7 \quad [k(ATT) \quad [m(ATT) \quad R'(ATT)]] \quad ]^{c'-c} >$ is the structured 

centered proposition corresponding to the clausal complement of

$ATT^7$).

To see that $\text{PRO}_s$, and hersel/her, are indeed Type-II covalued in (22), let us look at (25) – a 

more detailed version of (22).

---

We continue to assume the conventions in Footnotes 1, 2. Notice that in a theory where a description is not 

presupposed (see Footnote 6), this condition should be stated as follows.

Given a node $\chi$ dominated by $R(ATT)$ such that $[[\chi]]^{c'}$ is a ‘res’; for all $F \in C'_{NP1}^{c'}$ such that 

$F([[NP1]]^{c'})([[w(ATT)]]^{c'}) = [[\chi]]^{c'}$, for all $<w, x> \in \text{Dox}([[w(ATT)]]^{c'}, [[NP1]]^{c'})$, 

$$\{g'' \supseteq g': [[7 \quad [R'(ATT)]] \quad ]^{c'-c} \quad (F(x)(w)) \quad \text{is defined}\} \cap \{g'' \supseteq g'_{[k(ATT)-\omega]}: [[NP2]]^{c'-c} = x\} \neq \emptyset$$

where:

(i) $<[[\chi]]^{c'}, [[7 \quad [k(ATT) \quad [m(ATT) \quad R'(ATT)]] \quad ]^{c'-c} >$ is the structured centered proposition 

corresponding to the clausal complement of ATT; and

(ii) $R'(ATT)$ is just like $R(ATT)$ except that $t_7$ replaces $\chi$. 

---
ATT is \(\text{convince}^{\text{de-re}}\); \(w(\text{ATT})\) is \(w_0\) – the pronominal world-argument of \(\text{convince}^{\text{de-re}}\); \(S(\text{ATT})\) is the first node that dominates \(\text{convince}^{\text{de-re}}\) and all its syntactic arguments (i.e., \(\text{McCain}, w_0, t_6\), and the clausal complement of \(\text{convince}^{\text{de-re}}\) that is c-commanded by both \(k(\text{ATT})\) (namely, \(3\), the world-binding abstraction index introduced by \(\text{convince}^{\text{de-re}}\) and \(m(\text{ATT})\) (namely, \(5\), the individual-binding abstraction index introduced by \(\text{convince}^{\text{de-re}}\)). \(R(\text{ATT})\) is also the local domain of \(\text{herself/her}\). The reason why (24) renders \(\text{PRO}_5\) and \(\text{herself/her}_6\) Type-II covalued in (25) is the following.

Since according to (23), the clausal complement of \(\text{convince}^{\text{de-re}}\) must be defined in the doxastic alternatives of the ‘object’, the following holds for any relevant context \(c'\) such that \([\{c(\text{ATT}) \text{ McCain } \text{convince}^{\text{de-re}} \{w(\text{ATT}) \ w_0\} t_6 \ {l_k(\text{ATT})} \ 3 \ {l_m(\text{ATT})} \ 5 \ {l_{R(\text{ATT})/\text{LD}}} \ \text{PRO}_5 \ \text{vote}\ for-w_3 \ \text{herself/her}_6]]\] is defined and any relevant \(g' \supseteq g_c\) (in this case, just \(g_c\) itself): (a) any \(<w, x> \in \text{Dox(world}(c'))\), \([t_k]^{c', g'}\) is such that \([\text{PRO}_5 \ \text{vote}\ for-w_3 \ \text{herself/her}_6\] is defined (more accurately, \([\{7 \ [\text{PRO}_5 \ \text{vote}\ for-w_3 \ t_7]\] [\text{F}^{c, g'}(\text{world}(c')) = g'(6)]\) is defined, where \(\text{F}^{c, g'}(\text{world}(c')) = g'(6)\); (b) since \(\text{herself/her}_6\) and \(t_6\) are Type-I covalued, any \(<w, x> \in \text{Dox(world}(c'))\), \([\text{herself/her}_6]^{c', g'}\) is such that \([\text{PRO}_5 \ \text{vote}\ for-w_3 \ \text{herself/her}_6\] is defined, that is to say, for any \(<w, x> \in \text{Dox(world}(c'))\), \([\text{herself/her}_6]^{c', g'}\), \(\{g'' \supseteq g': [\text{PRO}_5 \ \text{vote}\ for-w_3 \ \text{herself/her}_6\] is defined} \cap \{g'' \supseteq g'_{[3-w]}\}; [\text{PRO}_5]^{c', g''} = x\} \neq \emptyset\) (more accurately, \(\{g'' \supseteq g': [\text{PRO}_5 \ \text{vote}\ for-w_3 \ \text{herself/her}_6\] is defined} \cap \{g'' \supseteq g'_{[3-w]}\}; [\text{PRO}_5]^{c', g''} = x\} \neq \emptyset\), where \(\text{F}^{c, g'}(\text{world}(c')) = g'(6)\). Therefore, by the definition of Type-II covaluation in
(24), *herself/her*₆ and *PRO₅* are Type-II covalued in (25), because the pair *<herself/her₆, PRO₅>* is such that for any relevant context *c′* such that \[[[S(ATT)McCain convince₃-w₃]]₆ \land [[(ATT)₃ \rightarrow [m₃(ATT)₅ \land (R(ATT)/LD₅ \land NP₂₅ PRO₅₃ vote for-w₃₃ [NP₁₆ herself/*her₆₆][])]]] \]

\[f" \supseteq g" \land \{g" \supseteq g': \text{}} \]] \[f"' \supseteq g"' \land \{g"' \supseteq g': \text{}} \]] \[f" \supseteq g" \land \{g" \supseteq g': \text{}} \]] \[f"' \supseteq g"' \land \{g"' \supseteq g': \text{}} \]] \[f" \supseteq g" \land \{g" \supseteq g': \text{}} \]] \[f"' \supseteq g"' \land \{g"' \supseteq g': \text{}} \]] \[f" \supseteq g" \land \{g" \supseteq g': \text{}} \]] \[f"' \supseteq g"' \land \{g"' \supseteq g': \text{}} \]] \[f" \supseteq g" \land \{g" \supseteq g': \text{}} \]] \[f"' \supseteq g"' \land \{g"' \supseteq g': \text{}} \]]

Notice that embedding *McCain convinced Palin to vote for herself/her* or replacing *Palin* with a quantifier won’t affect the predictions: *S(ATT)* is essentially unchanged in (26), so the reasoning that applies to (25) also applies here.

(26) [Obama believed-w₀ \[1 8 \text{no female politician-w₇₈}] \[6 \text{S(ATT)McCain convince₃-w₃}] \[w₆₃ \land [(\text{ATT})₃ \rightarrow [m₃(ATT)₅ \land (R(ATT)/LD₅ \land NP₂₅ PRO₅₃ vote for-w₃₃ [NP₁₆ herself/*her₆₆]])]]]]

It is also correctly predicted that replacing *herself/her* with *himself/him* (*McCain convinced Palin to vote for himself/him*) yields the desired result (namely, the non-reflexive variant is predicted to be acceptable, and the reflexive variant is predicted to be unacceptable), because *PRO* and *himself/him* are not Type-II covalued. The reason is this: the semantics of *convinces* does not require the ‘subject’ to believe the presuppositions of the clausal complement. Therefore, the set of possible contexts where the LF of this sentence is defined is not a subset of the set of possible contexts where McCain believes the presuppositions of the clausal complement (though there are, of course, many possible contexts where McCain does believe those presuppositions). To see this, let us look at (27).

(27) [McCain \[6 \text{S(ATT)} \land t₆ \text{convinces}_{w₃₆} \text{Palin} \[\text{[(ATT)}₃ \rightarrow [m₃(ATT)₅ \land (R(ATT)/LD₅ \land NP₂₅ PRO₅₃ vote for-w₃₃ [NP₁₆ himself/*him₆₆]])]]]]

*Him/him₆* and *PRO₅* are not Type-II covalued in (27), because there is at least one relevant
context \(c'\) such that \(\llbracket \llbracket_{\text{SATT}} t_6 \text{ convinced}^{\text{de-re}}_{\text{w(ATT)}} w_0 \rrbracket_{\text{Palin}} \rrbracket \llbracket_{\text{r(ATT)}} 3 \rrbracket_{\text{m(ATT)}} 3 \rrbracket_{\text{r(ATT)/LD}} [\text{NP2 PRO}_5 \text{ vote for-w}_3 \text{ [NP1 him/*hime}self]\]^{\text{w}(\text{ATT})}]\) \(\cup \llbracket \llbracket_{\text{SATT}} t_6 \text{ convinced}^{\text{de-re}}_{\text{w(ATT)}} w_0 \rrbracket_{\text{Palin}} \rrbracket \llbracket_{\text{r(ATT)}} 3 \rrbracket_{\text{m(ATT)}} 3 \rrbracket_{\text{r(ATT)/LD}} [\text{NP2 PRO}_5 \text{ vote for-w}_3 \text{ him/hi}msef]^{\text{w}(\text{ATT})}\) is defined, and yet there is some relevant \(g' \supseteq g_c\) and some \(<w, x> \in \text{Dox(world}(c'))\), \(g'(6)\) such that \(\{g'' \supseteq g': \llbracket \text{PRO}_5 \text{ vote for-w}_3 \text{ him/himsef}\]^{\text{w}(\text{ATT})} = x\} = \emptyset\) (similar results obtain if \(\text{PRO}_5\) is NP1 and \(\text{himself/hime}self\) is NP2).

To convince ourselves of this, let us consider context \(c_1\), where (i)-(iii) hold: (i) \(c_1\) supplies a function, \(F_{c_1} \in \llbracket_{\text{C1 Palin}} \rrbracket \), such that \(F_{c_1} \in \llbracket_{\text{Palin}} \rrbracket\)(world(\(c_1\))) = McCain and for all \(<w, x> \in \text{Dox(world}(c_1), \text{ Palin})\): \(\llbracket 7 \llbracket \text{PRO}_5 \text{ vote for-w}_3 t_7 \rrbracket \llbracket_{\text{m}(\text{ATT})} 3 \rrbracket_{\text{m}=\text{w}} F_{c_1}(w)(x)(w)\) is defined; (ii) \(<w_8, \text{ Obama}> \in \text{Dox(world}(c_1), \text{ McCain})\); and (iii) there is no \(F \in \llbracket_{\text{C1 McCain}} \rrbracket\) such that \(F(\text{McCain})(\text{world}(c_1)) = \text{McCain}\) and \(\llbracket 7 \llbracket \text{PRO}_5 \text{ vote for-w}_3 t_7 \rrbracket \llbracket_{\text{m}(\text{ATT})} 3 \rrbracket_{\text{m}=\text{w}} F_{c_1}(w)(x)(w)\) is defined, because the presuppositions of the clausal complement; whereas McCain entertains in world(\(c_1\)) the possibility that he is Obama, but doesn’t believe in world(\(c_1\)) all the presuppositions of the clausal complement. Still, \(\llbracket \llbracket_{\text{SATT}} t_6 \text{ convinced}^{\text{de-re}}_{\text{w(ATT)}} w_0 \rrbracket_{\text{Palin}} \rrbracket \llbracket_{\text{r(ATT)}} 3 \rrbracket_{\text{m(ATT)}} 3 \rrbracket_{\text{r(ATT)/LD}} [\text{NP2 PRO}_5 \text{ vote for-w}_3 \text{ him/himsef}]^{\text{w}(\text{ATT})}\) is defined, because the presuppositions of \(\text{convince}^{\text{de-re}}\) only require the ‘object’ to believe the presuppositions of the clausal complement; no such requirement holds of the ‘subject’.

9 In systems where \(\text{convince}^{\text{de-re}}\) is defined as in Footnote 6 and Type-II covaluation is defined as in Footnote 8, an example of a context where McCain doesn’t believe the presuppositions of the clausal complement (but Palin does) is c1-a, which looks like this: (i) \(\text{C1-a McCain} = \text{C1-a Palin}\), (ii) there is at least one \(F \in \llbracket_{\text{C1-a McCain}} \rrbracket\) such that \(F(\text{McCain})(\text{world}(c_1-a)) = F(\text{Palin})(\text{world}(c_1-a)) = \text{McCain}\), (iii) for all \(F \in \llbracket_{\text{C1-a Palin}} \rrbracket\), for all \(<w, x> \in \text{Dox(world}(c_1-a), \text{ Palin})\), \(\llbracket 7 \llbracket \text{PRO}_5 \text{ vote for-w}_3 t_7 \rrbracket \llbracket_{\text{m}(\text{ATT})} 3 \rrbracket_{\text{m}=\text{w}} F(x)(w)\) is defined and, (iv) at least one \(F \in \llbracket_{\text{C1-a McCain}} \rrbracket\) and one \(<w, x> \in \text{Dox(world}(c_1-a), \text{ McCain})\) are such that \(\llbracket 7 \llbracket \text{PRO}_5 \text{ vote for-w}_3 t_7 \rrbracket \llbracket_{\text{m}(\text{ATT})} 3 \rrbracket_{\text{m}=\text{w}} F(x)(w)\) is undefined. \(\llbracket \llbracket_{\text{SATT}} t_6 \text{ convinced}^{\text{de-re}}_{\text{w(ATT)}} w_0 \rrbracket_{\text{Palin}} \rrbracket \llbracket_{\text{r(ATT)}} 3 \rrbracket_{\text{m(ATT)}} 3 \rrbracket_{\text{r(ATT)/LD}} [\text{NP2 PRO}_5 \text{ vote for-w}_3 \text{ him/himsef}]^{\text{w}(\text{ATT})}\)
of c1: there is some \( g' \supseteq g_{c1} \) (namely, \( g_{c1} \)) such that for all \( \langle e,s,e' \rangle \)-functions \( F \in C1_{McCain}, F(McCain)(world(c1)) \neq McCain \) for some \( \langle w, x \rangle \in Dox(world(c1), McCain) \) (namely, \( \langle w8, Obama \rangle \)), \( \{ g'' \supseteq g'_{\{3\rightarrow w\}} : \lfloor PRO_5 \rfloor^{f_{\langle g' \rangle} = x} \cap \{ g'' \supseteq g''_{\{3\rightarrow w\}} : \lfloor PRO \rfloor_{\text{5 vote for-w_3}} \} \rfloor^{f_{\langle g' \rangle} (F(x)(w)) \text{ is defined}} = \emptyset \).

By contrast, \( \lfloor McCain convince^{de-re}_{\{w_{\text{ATT}}, w_0\}} t_6 \lfloor_\text{m(ATT) 3} \lfloor_\text{R(ATT)/LD PRO_5 vote for-w_3 herself/him} \rfloor_{\text{de-re}}^{\langle g' \rangle} \cap \text{the denotation of the relevant sub-LF of (25) relative to c2 – is undefined if c2 shares the pertinent properties of c1, namely: (i) } g_{c2}(6) = F_{c2,\text{R}(\langle 6 \rangle)(Palin)(world(c2)) = Palin, (ii) \langle w9, Obama \rangle \in Dox(world(c2), Palin); and (iii) \lfloor PRO_5 \text{ vote for-w_3}_t_7 \rfloor^{f_{\langle g' \rangle} (w) \text{ is defined and yet}) there is some relevant } g' \supseteq g_{c2} \text{ and some } \langle w, x \rangle \in Dox(world(c'), g'_{\lceil 6 \rfloor}) \text{ such that } \{ g'' \supseteq g': \lfloor PRO_5 \text{ vote for-w_3}_t_8 \text{ himself}_3 \rfloor^{f_{\langle g' \rangle} = x} \text{ is defined}} \cap \{ g'' \supseteq g'_{\{3\rightarrow w\}} : \lfloor PRO_5 \rfloor_{\text{5 vote for-w_3}} \} \rfloor^{f_{\langle g' \rangle} (F(x)(w)) \text{ is defined}} = \emptyset \text{. As a result, } PRO \text{ and himself/him are Type-II covalued. For the same reason, } McCain promised Palin to vote for himself \text{ is ruled out. But } McCain promised Palin to vote for herself \text{ is ruled out, once again because Palin is not required, by the}

\[ \lfloor^{1\text{A}_{\langle g' \rangle}} \text{ is defined and yet } PRO \text{ and him/himself fail to be Type-II covalued. This rendition of Type-II covaluation is compatible with the assumption that for any context } c \text{ and any } x, C_{\text{c}} \text{ includes } [\lambda x \in D_{\text{c}} . \lambda w \in D_{\text{c}} . x] \).

10 We are assuming (but see references in Footnote 3) that the ‘subject’ of convince may do the actual convincing indirectly, even accidentally, without uttering the clausal complement or even having in mind a thought that is represented by the clausal complement (e.g., The movie convinced me to vote for Palin, although it supported Obama). But even if we said that the ‘subject’ is required, after all, to entertain the presuppositions of the clausal complement, this would be true in (27) only in the ‘de te’ sense, not the ‘de se’ sense (see Appendix); therefore we would still predict McCain convinced Palin to vote for himself to be unacceptable and McCain convinced Palin to vote for him to be acceptable.
semantics of promise, to believe the presuppositions of the clausal complement (so there are contexts where she indeed doesn’t, yet the relevant LF is still defined; as a result, PRO and herself are not Type-II covalued).

Next, McCain expects Palin to vote for himself, where the embedded subject is Palin, is ruled out, and its non-reflexive variant is ruled in. This variety of expect is not a Control verb, so for every y ∈ De, including McCain, there are contexts where y doesn’t take herself/himself to be Palin (i.e., there are relevant contexts c’ and assignments g’ ⊇ g_c where for some <w, x> ∈ Dox(world(c’), g’(6)), {g” ⊇ g’[3→w]: [[Palin]]f’g” = x} = ∅, so it follows that {g” ⊇ g’: [[Palin vote for-w3 himself/him]]f’g” is defined} ∩ {g” ⊇ g’[3→w]: [[Palin]]f’g” = x} = ∅). Consequently, Palin and himself/him are not Type-II covalued.

The readers are referred to the appendix for some necessary refinements of the theory, and discussion of the cases that motivate those refinements (namely, verbs that support ‘de te’ readings of PRO). The next section discusses further consequences of the proposal: De Re and De Se Blocking, Free Indirect Discourse, and Exceptional Case Marking constructions.

4. Further consequences

4.1. De Re Blocking

The term De Re Blocking refers to the inability of a reflexive pronoun that is c-commanded by a ‘de re’ pronoun to be interpreted ‘de se’ (as first observed in Sharvit 2009, and later elaborated on in Charlow 2010a,b). Consider the sentences corresponding to McCain convinced Palin to vote for herself and McCain promised Palin to vote for himself, where both embedded pronouns are overt, as in (28). The reflexives in these sentences are ambiguous in a way that parallels their ambiguity in their Control counterparts.11

(28) a. McCain convinced Palin that she (had) voted for herself.
   Palin’s conclusion: “I voted for myself”.
   Palin’s conclusion: “I voted for this woman”.

b. McCain suggested to Palin that he (had) voted for himself.
   McCain’s suggestion: “I voted for myself”.

11 We are ignoring the semantic differences between Control convince and non-Control convince. The only difference relevant to the current discussion is that the latter has an overt embedded subject.
McCain’s suggestion: “I voted for this guy.”

Suppose Palin doesn’t remember who she voted for. McCain can convince her that she voted for her ‘self’; or for Palin. Similarly, McCain can suggest to Palin that he voted for his ‘self’; or for McCain. Interestingly, what (28a) cannot mean is that Palin’s conclusion is “This woman voted for me” (though it can be “This woman voted for herself”). Likewise, (28b) cannot mean that McCain suggestion was “This guy voted for me” (though it could be “This guy voted for himself”).

This De Re Blocking effect occurs only when the embedded subject is overt; the covert PRO is an inherent ‘de se’ pronoun, as shown by the well-known contrasts in (29) and (30).

(29)  
a. Obama expects/hopes that he will get re-elected.  
b. Obama expects/hopes to get re-elected.

(30)  
 a. McCain convinced Palin that she should retire.  
b. McCain convinced Palin to retire.

While (29a) is ambiguous between a ‘de re’ reading (which reports a ‘de re’-expectation/hope: “This guy will get re-elected”) and a ‘de se’ reading (which reports a ‘de se’-expectation/hope: “I will get re-elected”), the Subject-Control construction (29b) has only a ‘de se’ reading (Morgan 1970, Chierchia 1989). This is evidenced by the fact that (29b) cannot be uttered in a situation where Obama is looking at his own picture, failing to recognize himself in the picture and saying to himself: “This guy will get re-elected”, whereas (29a) can. In Chierchia (1989), these facts are taken to mean that the subject of the embedded clause in (29b) – PRO – is an inherent ‘de se’ pronoun, whereas he in (29a) is not (it can be understood ‘de re’ or ‘de se’). Likewise, while (30a) is ambiguous between a ‘de re’ reading (which reports Palin’s ‘de re’ decision: “This woman should retire”) and a ‘de se’ reading (which reports Palin’s ‘de se’-decision: “I should/will retire”), (30b) has only a ‘de se’ reading. Consistent with Chierchia’s analysis of PRO, McCain convinced Palin to vote for herself doesn’t have a reading that implies that Palin decided, pointing at her own picture, “This woman should vote for herself.”

As the readers have probably noticed, our proposal, as stated in Section 3, is incompatible with the facts. While we can account for the readings in (28) by assuming that the overt subject
pronoun may be bound by convince/suggest (and thus be Type-II covalued with the reflexive), we also predict the reading that does not exist, namely, the reading where the reflexive is ‘de se’ and the overt pronominal subject is a ‘de re’ pronoun: in this case, too, the subject and the object are Type-II covalued.

To avoid this unwelcome prediction, we restate Condition A as follows: a reflexive pronoun must be covalued with some suitable syntactic variable in it local domain (where a ‘suitable syntactic variable’ is a trace or a pronoun bound directly by an abstraction index, without the intervention of another c-commanding trace/pronoun). Thus, (31) and (33), where the reflexive is either Type-I or Type-II covalued with a suitable local syntactic variable (\(t_2\) and \(she/PRO\) respectively), obey the new Condition A, but (32), and also (34) (the LF underlying the missing reading of (28a)), do not.

(31) \([Palin/She \[2 \[LD t_2 \text{ vote for herself}_2]]\]
(32) \(*[LD Palin/She_{\text{o}} \text{ vote for herself}_2]\)
     \((she_{\text{o}} \text{ is not a suitable syntactic variable because it is free}; Palin \text{ is not a variable})\)
(33) \([Palin \[6 \[McCain convinced^{de-re-w_0} t_6 \[3 5 \[LD she/PRO_{\text{o}} \text{ vote for-w_3 herself}_2]]]]\]
(34) \(*[Palin \[6 \[McCain convinced^{de-re-w_0} t_6 \[3 5 \[LD she_{\text{o}} \text{ vote for-w_3 herself}_2]]]]\)
     \((she_{\text{o}} \text{ is not a suitable syntactic variable because } t_6 \text{ intervenes between 6 and } she_{\text{o}})\)

It is worth noting that the De Re Blocking effect that reflexive pronouns exhibit is reminiscent of De Re Blocking effects with non-reflexive pronouns (first reported in Percus & Sauerland 2003b), as illustrated by the dream report in (35).

(35) I dreamed I was Jesus and I forgave my mother.

Possible interpretation: In the dream, Jesus forgives dreamer’s mother.

Impossible interpretation: In the dream, dreamer forgives Jesus’s mother.

But importantly, reflexive and non-reflexive pronouns do not exhibit the same De Re Blocking pattern. As noted by Anand (2006) and by Percus & Sauerland, a ‘de se’ reading of a non-reflexive pronoun is not blocked by a c-commanding ‘de re’ pronoun when the attitude verb is a believe/convince/suggest-type verb (i.e., when the report is not a dream report). Crucially,
Anand’s observation applies to non-reflexive pronouns only. Reflexive pronouns show a
different pattern: *dream* and *convince* alike do not allow a ‘de se’ interpretation of a reflexive
pronoun c-commanded by a ‘de re’ pronoun, as shown by (28) above, and by (36) (from Heim
1994a, where unexpected Condition A were first observed). On the other hand, regarding non-
reflexive pronouns, *convince* and *suggest* do not behave like *dream* (in (35)): there is no De Re
Blocking when the second pronoun is non-reflexive, as (37) shows. Thus, the patterns in (38)
seem to emerge.

(36) I dreamed I was Jesus and I forgave myself.
   Possible interpretation: In the dream, Jesus forgives dreamer.
   Impossible interpretation: In the dream, dreamer forgives Jesus.

(37) a. McCain convinced Palin that she (had) voted for her opponent.
   Possible interpretation – Palin’s conclusion: “This woman voted for my opponent.”
   Context: Palin lost her memory after the elections. McCain shows Palin a picture of
   herself, but she fails to recognize herself in the picture. He says to Palin: “You may
   not believe it, but this woman voted for your opponent.” Palin believes him.

b. McCain suggested to Palin that he (had) voted for his opponent.
   Possible interpretation – McCain to Palin: “This guy voted for my opponent.”
   Context: McCain lost his memory after the elections. He is looking, together with
   Palin, at a picture of himself, but fails to recognize himself in the picture. He says: “I
   have a nagging feeling that this guy voted for my opponent.”

(38) \[
\begin{array}{ll}
\text{Reflexive pron}^{\text{de se}} & \text{Non-reflexive pron}^{\text{de se}} \\
\text{dream} & \neq \left[\ldots\text{pron}^{\text{de re}} \ldots \ldots \ldots \text{pron}^{\text{de se}} \ldots\right] \quad (36) & \text{[]} \left[\ldots\text{pron}^{\text{de re}} \ldots \ldots \ldots \text{pron}^{\text{de se}} \ldots\right] \quad (35) \\
\text{convince} & \neq \left[\ldots\text{pron}^{\text{de re}} \ldots \ldots \ldots \text{pron}^{\text{de se}} \ldots\right] \quad (28) & \checkmark \left[\ldots\text{pron}^{\text{de re}} \ldots \ldots \ldots \text{pron}^{\text{de se}} \ldots\right] \quad (37)
\end{array}
\]

This paper has nothing to say about De Re Blocking in dream reports (and the readers are
referred to Anand 2006 and references cited there for possible explanations). Here we are
concerned solely with De Re Blocking with reflexives, which is accounted for with the new
Condition A. Since we ignore dream reports here, we take non-reflexive pronouns to show no De
Re Blocking.

Finally, as we saw (in Section 2), and as predicted by the theory presented in Section 3,
non-reflexive ‘de re’ pronouns respect the new Condition B (“a non-reflexive pronoun may not
be covalued with any c-commanding NP in its local domain,” where ‘covalued’ is understood as ‘Type-I or Type-II covalued’). As shown by (39) (and as also predicted by the new Condition B), non-reflexive ‘de se’ pronouns do too.

(39) McCain visits three amnesiac female politicians, in order to talk to them about the recent elections.

a. He must have convinced every amnesiac female politician that she had voted for herself/*her, because each one of them is now holding a picture of herself saying: “I voted for this woman.”

Relevant reading: For all female politician x, McCain convinced x that x’s ‘self’ voted for x.

b. *He must have convinced every amnesiac female politician that she had vote for herself/her, because each one of them is now holding a picture of herself saying: “This woman voted for me.”

Relevant reading: For all female politician x, McCain convinced x that x voted for x’s ‘self’.

(39a) is acceptable with a reflexive ‘de re’ pronoun (which is Type-II covalued with she, in satisfaction of Condition A) but unacceptable with a non-reflexive ‘de re’ pronoun (which violates Condition B). (39b) is simply unacceptable: a reflexive ‘de se’ pronoun violates the new Condition A (because the reflexive is indeed covalued with she, but she is not a suitable syntactic variable) and a non-reflexive ‘de se’ pronoun is Type-II covalued with she, in violation of the new Condition B.

Charlow (2010a) shows that ‘de se’ pronouns can obviate Condition B. Indeed, non-reflexive ‘de se’ pronouns obviate Condition B in special circumstances (provided by an appropriate context), but so do ‘de re’ pronouns (see discussion of (17) and (18) in Section 2). Improved versions of (39) with non-reflexive pronouns are give in (40).

(40) a. McCain showed every amnesiac female politician a picture of a woman, namely, herself. He must have convinced every politician that she had voted for her; because each one of them is now saying: “I voted for this woman.”
b. Yesterday, all the amnesiac female politicians where wondering: “Who on earth could have voted for me?” McCain must have convinced every female politician that SHE (HERSELF) had voted for her, because each one of them is now holding a picture of herself mumbling: “This woman voted for me.”

In the current theory, the improved status of *McCain convinced every female politician that she had voted for her* in special circumstances where she is understood ‘de se’ and her ‘de re’, as well as its improved status in special circumstances where she is understood ‘de re’ and her ‘de se’, are expected to be accounted for by a single theory that predicts obviation of Condition B in special circumstances (whatever that theory might be). Importantly, (39a,b) are uttered in contexts where obviation of Condition B seems impossible (or extremely hard).

### 4.2. De Se Blocking

The term De Se Blocking refers to the inability of a reflexive pronoun that is c-commanded by a ‘de se’ pronoun – in VP-ellipsis constructions and *only*-constructions – to be interpreted ‘de re’. This phenomenon is fairly complex, so to set the stage, we first discuss the obligatory “sloppy” reading of reflexives, and then see how it relates to De Se Blocking.

Consider the VP-ellipsis in (41a), and the *only*-construction in (41b).

(41) a. McCain convinced Palin that she had voted for herself while no other female politician had.

b. McCain convinced Palin that only she had voted for herself.

The most salient bound-variable interpretations of (41a,b) are those represented by the LF in (42), according to which Palin’s conviction is: “I am/this woman is the only one with the property \[ \lambda x \in D_e . x \text{ voted for } x \].”

(42) \[ [...[3 5 [[she_{65} [2 [t_2 \text{ vote for-}_w3 \text{ herself}_2]]]] while [no other politician-}_w30 [2 [t_2 \text{ vote for-}_w3 \text{ herself}_2]]]]]]]
But (41a,b) have another bound-variable interpretation – a dual ‘de re’ bound-variable interpretation, where both embedded pronouns – *she* and *herself* – are interpreted ‘de re’ and ‘bound’. The scenario described in (43) is appropriate for such a reading. The dual (bound) ‘de re’ reading cannot be accounted for by the LF in (42).

(43) Palin is looking at pairs of pictures of politicians: two pictures of herself, and two pictures of each of the other female politicians. She doesn’t recognize herself in either one of her pictures. In addition, for each pair of pictures, she fails to recognize that the two pictures are of the same person. In other words, she thinks that the two pictures of herself are of Candidate A and Candidate A’, that the two pictures of Diane Feinstein are of Candidate B and Candidate B’, and the two pictures of Caroline Kennedy are of Candidate C and Candidate C’. As a result of talking to McCain, she reaches the following conclusion: “Candidate A voted for Candidate A’, while Candidate B didn’t vote for Candidate B’ and Candidate C didn’t vote for Candidate C’.”

What is remarkable about this interpretation is that every pair may be “counted” once (that is to say, Palin is not required to care whether B’ voted for B), so that we may paraphrase Palin’s belief as: “Of the pairs {<A, A’>, <B, B’>, <C, C’>}, only the pair <A, A’> is such that the first member voted for the second member.” (41a,b) also have a reading that can be paraphrased as: “Of the pairs {<A, A’>, <B, B’>, <C, C’>}, only the pair <A, A’> is such that at least one member voted for the other,” but the point about De Se Blocking can be made without it.)

The dual bound ‘de re’ interpretation contrasts with the interpretation which, if it were available, would be appropriate in the scenario in (44), which requires both embedded pronouns to be interpreted ‘de re’, but allows only the subject to be ‘bound’. It also contrasts with the interpretation which, if it were available, would be appropriate in the scenario in (45), which allows only the second pronoun to be interpreted ‘de re’.

---

12 In Sharvit (2009), I claimed that *she* and *herself* cannot both be ‘de re’ in *McCain convinced Palin that she voted for herself*, even when two different descriptions are used; but Simon Charlow has pointed out to me that this reading is available, and that stress on *herself* facilitates it. What I discovered since is that stress on *herself* also facilitates a bound variable dual ‘de re’ reading.
(44) Palin is looking at a pair of pictures of herself, failing to recognize herself, and failing to recognize that the pictures are of the same person: she thinks that one picture is of Candidate A and the other picture is of Candidate B. After talking to McCain, she says, pointing at the first picture: “Candidate A voted…” and then pointing at the second picture: “… for Candidate B, but none of those other people voted for Candidate B.”

(45) Palin is looking at a picture of herself, failing to recognize herself: she thinks that the picture is of Candidate B. After talking to McCain, she points at the picture and says: “I voted for Candidate B, but none of those other people voted for Candidate B.”

While several speakers accept (41a,b) in the scenarios described in (44) and (45), it seems that for the majority of the speakers, the reading corresponding to (43) – where herself is bound – is the only possible ‘de re’ reading. This is reminiscent of the fact that quite generally, herself supports only “sloppy” readings. For example: John scratched himself but Bill didn’t implies that Bill didn’t scratch Bill (cf. John scratched his foot but Bill didn’t with a non-reflexive pronoun, which is ambiguous between a “strict” and “sloppy” reading, and need not imply that Bill didn’t scratch Bill’s foot). But even here, there seem to be exceptions among speakers (for discussion and concrete proposals, see Sag 1976 and Hestvik 1995, among others). The contrast represented by (43)-(45) is of the same nature: most speakers accept sloppy ‘de re’ readings of (41a,b) and do not accept strict ‘de re’ readings at all.

Importantly, there are two more types of scenarios where (41a,b) are inappropriate. These scenarios resemble the one in (43) in that Palin looks at pairs of pictures of all the other candidates and doesn’t acknowledge, for each pair, that the two pictures are of the same person. But unlike the situation in (43), she looks at only one picture of herself (without recognizing herself in the picture). It is impossible to interpret (41a,b) as if Palin reaches either one of the following conclusions: (i) “Candidate A voted for me, but Candidate B didn’t vote for Candidate B’, and Candidate C didn’t vote for Candidate C’,” (ii) “I voted for Candidate A, but Candidate B didn’t vote for Candidate B’ and Candidate C didn’t vote for Candidate C’.” That (i) is impossible is perhaps not unexpected, given that reflexives exhibit De Re Blocking. That (ii) is impossible shows that reflexives also exhibit De Se Blocking (see also (45)): In VP-ellipsis/only constructions of the type illustrated in (41), a reflexive cannot be interpreted ‘de re’ if c-commanded by a ‘de se’ pronoun.
How can we account for De Se Blocking of reflexives? First, we suggest that the LF in (46) is the LF of the embedded clause of (41), on the dual bound-variable ‘de re’ reading of (41) (corresponding to the scenario described in (43)).

(46) \[3 5 [ [[\textit{she}_7 6 [t_6 \textit{[herself]_6]]] [2 4 [t_4 \textit{vote for-w}_3 t_2]]] [\textit{while} [no^\Omega [\textit{polit-w}_0 [\textbf{OP} [6 [t_6 \textit{[herself]_6]]]] [2 4 [t_4 \textit{vote for-w}_3 t_2]]]]]

This LF is “unorthodox” in at least two respects: (i) \(t_4\) and \(t_2\) are traces of \(\textit{she}_7\) and \(\textit{herself}\), yet they are not c-commanded by them; and (ii) the elided parts are \([6 [t_6 \textit{[and] [herself]_6]]] [2 4 [t_4 \textit{vote for-w}_3 t_2]]\), which do not form a constituent. Both (i) and (ii) are at odds with standard assumptions about ellipsis. The relevant reading is derived on the following assumptions: (a) \([\textit{she}_7 6 [t_6 \textit{[herself]_6]]]\) denotes a pair of individuals, (b) \([\textit{politician-w}_0 [\textbf{OP} [6 [t_6 \textit{[and] [herself]_6]]]]]\) denotes a set of such pairs (and the context, in principle, supplies as many descriptions as needed), (c) \([2 4 [t_4 \textit{vote for-w}_3 t_2]]\) denotes a function from pairs of individuals to truth values, (d) \(\textbf{OP}\) yields a set of pairs, and (e) \(no^\Omega\) is a quantifier over pairs, as shown in (47).\(^{13}\)

\[
\begin{align*}
([t_6 \textit{[and] [herself]_6]]) & \xrightarrow{\text{\textbf{C}G}[6=x]} = \langle x, x \rangle \\
([6 [t_6 \textit{[herself]_6]]]) & \xrightarrow{\text{\textbf{C}G}} = [\lambda x \in D_e. \langle x, x \rangle] \\
([\textit{she}_7 6 [t_6 \textit{[herself]_6]]]) & \xrightarrow{\text{\textbf{C}G}} = \langle[[\textit{she}_7]\text{\textbf{G}},[[\textit{she}_7]\text{\textbf{G}}]

[\textbf{OP}]\xrightarrow{\text{\textbf{C}G}} = [\lambda f \in D_{\textit{ex,x}}: \text{for all } y \in \text{Dom}(f), f(y) \text{ is an ordered pair} . \lambda X \in D_e : X \text{ is an ordered pair such that there is a } y \in D_e \text{ such that } \text{First}(X) = \text{First}(f(y)) \text{ and Second}(X) = \text{Second}(f(y)) . \text{P(First}(X)) \text{ and P(Second}(X)) \text{] (where for any pair } <a, b> : \text{First}(<a, b>) = a \text{ and Second}(<a, b>) = b). \\
[\textit{polit-w}_0 [\textbf{OP} [6 [t_6 \textit{[and] [herself]_6]]]]] & \xrightarrow{\text{\textbf{C}G}} = [\lambda X \in D_e : \text{there is a } y \in D_e \text{ such that } \text{First}(X) = \text{Second}(X) = y . \text{First}(X) \text{ is a female politician in world(c) and}

\]

\(^{13}\) We also have to assume a type-shifting operation that turns \([\text{\textit{she}_7 6 [t_6 \textit{[herself]_6]]}]\) \xrightarrow{\text{\textbf{G}}} \in \lambda R \in D_{\textit{ex,ex}} . R(\text{Second}([\text{\textit{she}_7 6 [t_6 \textit{[herself]_6]]}]) \xrightarrow{\text{\textbf{G}}}) (\text{First}([\text{\textit{she}_7 6 [t_6 \textit{[herself]_6]]}]) \xrightarrow{\text{\textbf{G}}}) = True). In addition, \textit{no politician} is actually \textit{no other politician}, and for every \(x\), \textit{no other(x) politician} yields a set of pairs that excludes \(\langle x, x \rangle\).
Second(X) is a female politician in world(c)

\[[\text{no}^\Omega]^\text{fg} = [\lambda f \in D_{\text{fg},D^c} : \text{for all } Y \in \text{Dom}(f), Y \text{ is an ordered pair. } \lambda R \in D_{\text{fg},D^c} \cdot \{Y : f(Y) = \text{True}\} \cap \{Y : R(\text{Second}(Y))(\text{First}(Y)) = \text{True}\} = \emptyset]\]

Whenever defined (i.e., at the very least, whenever:

(i) \(F_1^\text{Palin}(\text{Palin})(\text{world}(c)) = F_2^\text{Palin}(\text{Palin})(\text{world}(c)) = \text{Palin}\), and

(ii) for all \(y \neq \text{Palin}\) such that \(y\) is a female politician in world(c),

\(F_1^y(\text{Palin})(\text{world}(c)) = F_2^y(\text{Palin})(\text{world}(c)) = y\),

\[[\text{Palin}[7 [\text{McCain convinced}^\text{de-re-w}_0 t_7 [(46)]]]]^\text{fgc} = \text{True iff for all } <w, x> \in \text{Dox(\text{world}(c), \text{Palin})}: F_1^\text{Palin}(x)(w) \text{ voted for } F_2^\text{Palin}(x)(w) \text{ in } w, \text{ and for all } y \neq \text{Palin} \text{ such that } y \text{ is a female politician in } \text{world}(c), F_1^y(x)(w) \text{ didn’t vote for } F_2^y(x)(w) \text{ in } w.\]

This LF obeys Condition A, because the reflexive is locally bound by a trace (a suitable syntactic variable). In fact, we are forced to say that whenever both \(\text{she}\) and \(\text{herself}\) are ‘de re’ (even without \(\text{only}\) or VP-ellipsis, as in \(\text{McCain convinced Palin that she voted for herself}\)), the underlying LF is “unorthodox” (as in (48a)); otherwise, Condition A is not met: in (48b), \(\text{herself}\) is not covalued, in any sense, with a c-commanding suitable syntactic variable (cf. (32)).

(48) a. \(\ldots [3 \ 5 [[[\text{she}^\gamma_7 \ 6 \ \text{[LD } t_6\text{]] [[[\text{herself}^\gamma_6]]]]]] [(46)]] [2 \ 4 \ [t_4 \ \text{vote for-w}_3 t_2]]]]\]

b. *\(\ldots [3 \ 5 \ \text{[LD she}^\alpha_6 \ \text{voted for-w}_3 \ \text{herself}^\gamma_6]]]\)

Importantly, on its own, the new Condition A does not suffice to account for either De Re or De Se Blocking. This is because, in principle, one of the members in the pair of descriptions used to interpret \([\text{she}^\gamma_7 \ 6 \ \text{[LD } t_6\text{] [\text{herself}^\gamma_6]]]}\) in (46)/(48a) may be the identity function/description (i.e., \(\lambda x \in D_e . \lambda w \in D_s . x\)), which effectively yields a ‘de se’ interpretation. In point of fact, it seems that ‘de re’ pronouns are indeed compatible with ‘de se’ situations, as shown by examples with bound pronouns of the sort discussed in Zimmermann (1991). Consider Every candidate, including George, expects that he will get re-elected, which is

\(^{14}(46)\) cannot make (41) felicitous in either (44) or (45), because in those scenarios Palin recognizes that the other relevant individuals are distinct from each other and distinct from her.
felicitous in a situation where George’s expectation, when pointing at his own picture (and failing to recognize himself in the picture) is: “This guy should get re-elected,” while all the other candidates think: “I should get re-elected.” This is easily accounted for if he is taken to be a bound ‘de re’ pronoun, and the identity function is considered a suitable description for ‘de re’ expressions (see (23)).

(49) \[
\text{[every candidate-w}_0 \ [7 \ [t_7 \ \text{expects}^{\text{de-re-w}_0} \ [3 \ 4 \ [\text{he}_7 \ \text{will-w}_3 \ \text{get \ re-elected}]])}\]

But if the identity function is indeed a suitable description in principle, this means that one of the members of the pair of descriptions used to interpret \[
\text{[she}_6 \ [6 \ [\text{LD} \ [t_6] \ [\text{herself}_6])]]\]
in (46)/(48a) could be the identity function. This, in turn, means that De Se Blocking is not predicted at all. It also means that De Re Blocking may be overridden, because (46)/(48a) obey Condition A. The point about overriding De Re Blocking is made somewhat differently in Charlow (2010a,b), since for Charlow an LF such as (48b), where a ‘de re’ reflexive is c-commanded by a ‘de re’ non-reflexive pronoun, is not ungrammatical. Charlow notes that allowing herself to be interpreted relative to the identity function indeed predicts, incorrectly, the circumvention of De Re Blocking (but disallowing the use of the identity function might face problems in view of Zimmermann-type examples).

I would therefore like to claim that just like the Zimmermann-type example suggests that the identity function is, in principle, a suitable description for the interpretation of ‘de re’ expressions, De Se Blocking suggests something quite different, namely, that the preferred ‘de se’ mechanism is binding of a pronoun by an attitude verb. Accordingly, De Se Blocking is expected in (50), where she is bound by the matrix verb (so \[
\text{[she}_5 \ [6 \ [\text{LD} \ [t_6] \ [\text{herself}_6])]]\]
is interpreted ‘de se’, as <‘self’, ‘self’>; that is to say, as <x, x> for any relevant doxastic alternative <w, x>).

(50) \[
\ldots \ [3 \ 5 \ [[\text{[she}_5 \ [6 \ [\text{LD} \ [t_6] \ [\text{herself}_6]]]]] \ [2 \ 4 \ [t_4 \ \text{vote \ for-w}_3 \ \text{t}_2]]\ldots]]\]

\[\overline{15}\]

For some (though admittedly not all) speakers, (41) cannot imply that Palin decided: “I voted for myself while Candidate B didn’t vote for Candidate B’ and Candidate C didn’t vote for Candidate C.” This can be explained in terms of Economy (along the lines of Fox 2000): the LF in (50) is overridden by \[
\ldots \ [3 \ 5 \ [[\text{she}_5 \ [2 \ [\text{LD} \ [t_2] \ \text{vote \ for-w}_3 \ \text{herself}_5]]]]\ldots]]\]
where the sub-tree of the first conjunct is simpler.

\[\overline{15}\]
In other words, De Se Blocking suggests, at the very least, that the use of the identity function is restricted. What exactly the relevant restrictions are is a difficult question (the answer to which is beyond the scope of this paper). Whatever these restrictions are, they must be such that even if they allow the use of the identity function in “orthodox” LFs (e.g., (49)), they don’t always allow it in “unorthodox” LFs (e.g., (46)/(48a)). This conclusion is confirmed by the fact that (51b) is not a possible interpretation of (51a). The reason is that, for some (yet unknown) reason, the identity function cannot be used to interpret any of the ‘de re’ expressions in (52), which contains the “unorthodox” LF in (46).

(51) a. McCain convinced every female politician that only she had voted for herself.
   b. Some female politician (looking at a picture of herself, not recognizing herself in it, and at pairs of pictures of the other politicians) says: “This candidate voted for me but Candidate B didn’t vote for Candidate B’ and Candidate C didn’t vote for Candidate C,’” while another female politician (looking at a picture of herself, not recognizing herself in it, and at pairs of pictures of the other female politicians) says: “I voted for this candidate, but Candidate B didn’t vote for Candidate B’, and Candidate D didn’t vote for Candidate D’.”

(52) [every politician-w_0 [7 [McCain convinced_{de-re-w_0} t, [(46)]]]]

More to the point, De Se Blocking shows that the grammar allows binding of pronouns by attitude verbs (as in (50)). This means that while some pronouns that are understood ‘de se’ may, perhaps, be ‘de re’ in the syntactic sense, the grammar definitely also generates pronouns that are ‘de se’ in the semantic as well as the syntactic sense. We come back to this point in Section 5.

Finally, recall from Section 4.1 that non-reflexive pronouns do not exhibit the same De Re Blocking effects as reflexive pronouns. As is well known (and mentioned above), non-reflexive pronouns are also different from reflexive pronouns in that they support strict readings (for example, John scratched his foot and Bill did too need not imply that Bill scratched his own foot). Thus (53a), unlike (41a,b), can imply that Palin concluded (while pointing at a picture of herself, thinking it is a picture of Candidate B): “I this woman voted for the opponent of Candidate B, while none of the other politicians voted for the opponent of Candidate B.” This is
predicted by the “orthodox” LF in (53b) (an “unorthodox” LF is also available, in principle\textsuperscript{16}).

(53) a. McCain convinced Palin that she had voted for her opponent while no other female politician had.

  b. \[\text{Palin} \; [6 \; [\text{McCain convinced-w_0} \; t_6 \; [3 \; 5 \; [\text{she}_6/5 \; [2 \; t_2 \; \text{vote for-w_3} \; [\text{LD} \; \text{her}_6 \; \text{opponent-w_3}]]] \; \text{while no other politician-w_0} \; [2 \; t_2 \; \text{vote for-w_3} \; [\text{LD} \; \text{her}_6 \; \text{opponent-w_3}]]]]] \]

An “orthodox” LF cannot, of course, be used in the case of (41), as this would lead to a violation of Condition A.

The contrasts between reflexive and non-reflexive pronouns regarding the availability of strict/sloppy readings (and regarding De Re Blocking, ignoring dream-type verbs) are sketched in (54).

(54) a. \ldots[3 \; 5 \; [[\text{she}_6/5 \; [2 \; t_2 \; \text{vote for-w_3} \; \text{her}_7 \; \text{opponent-w_3/\s{herself}_3}]]] \ldots

  b. \ldots[3 \; 5 \; [[\text{she}_5 \; \text{vote for-w_3} \; \text{her}_5/6 \; \text{opponent-w_3/\s{herself}_5/6}]]] \ldots

  c. \ldots[3 \; 5 \; [[\text{she}_6/5 \; [2 \; t_2 \; \text{vote for-w_3} \; \text{her}_6/5 \; \text{opponent-w_3/\s{herself}_6/5}]]] \ldots

  d. \ldots[3 \; 5 \; [[\text{she}_6 \; \text{vote for-w_3} \; \text{her}_5/6 \; \text{opponent-w_3/\s{herself}_5/6}]]] \ldots

  e. \ldots[3 \; 5 \; [[[\text{she}_{7/5} \; [6 \; [[\text{her}_6 \; \text{opponent-w_3/\s{herself}_3}]]] \; [2 \; 4 \; t_4 \; \text{vote for-w_3} \; t_2]]] \ldots

  f. \ldots[3 \; 5 \; [[[\text{she}_{7/5} \; [6 \; [[\text{her}_7 \; \text{opponent-w_3/\s{herself}_3}]]] \; [2 \; 4 \; t_4 \; \text{vote for-w_3} \; t_2]]] \ldots

It is also worth noting that the requirement that reflexives be locally covalued with a suitable syntactic variable is needed only to account for De Re Blocking; it is not needed to account for the absence of strict readings (though it is compatible with the absence of strict readings).

4.3. Unexpected BT effects in Exceptional Case Marking constructions

As observed in Chierchia (1989), a reflexive pronoun in the subject position of a complement of an E(xceptional) C(ase) M(arking) verb is ambiguous between a ‘de re’ and a ‘de se’ reading.

\textsuperscript{16} The availability of an “unorthodox” LF allows us to correctly predict a dual bound ‘de re’ reading for (53a). Together with that, we predict non-reflexive pronouns to also show De Se Blocking. That is to say, we predict that (53a) does not have a reading according to which Palin decides: “I voted for the opponent of Candidate A, but Candidate B didn’t vote for the opponent of Candidate B’, and Candidate C didn’t vote for the opponent of Candidate C.” Speakers judgments are too non-uniform for us to decide whether the latter is a welcome prediction.
(55) Joe Biden wants himself to be the next president.

That is to say, (55) is compatible with a situation where Joe Biden points at a picture of himself, and says “I wish this guy were the next president”, as well as a situation where he says “I wish I were the next president.”\(^\text{17}\)

As we already know, a ‘de re’ LF can account for the ‘de se’ reading, provided the identity function counts as a suitable description for a ‘de re’ pronoun. Thus, a ‘de se’ interpretation is of *himself* may be obtained from the ‘de re’ LF in (56), when the relevant description happens to be an identity function (assume, for simplicity, that *want* quantifies over the doxastic alternatives of the ‘subject’; the bold brackets mark the local domain of the embedded subject reflexive, which in ECM constructions does not correspond to any node in the tree).

(56) \([Biden\ [5 \lf LD t_5 \text{wants}^{de-re} w_0 \lf 3 \lf 4 \lf [\text{himself}_s] \text{to be the next president-w}_3)]])\]

The local domain of the embedded subject includes the ECM verb and the matrix subject. Importantly, *himself* and *t* in (56) are Type-I covalued, and *himself* is locally bound by the trace of *Biden*; therefore, (56) satisfies Condition A via Type-I covaluation.

But ECM constructions exhibit unexpected Condition B effects, as evidenced by the fact that (57) doesn’t have a bound variable ‘de se’ reading.

(57) Every candidate wants him to be the next president.

As long as we assume that ‘de se’ pronouns may be syntactically ‘de se’ (i.e., bound by the verb), or syntactically ‘de re’ (via the identity function), a bound variable ‘de se’ reading can, in principle, arise from both LFs in (58).

(58) a. \([\text{every candidate} \lf 2 \lf [LD t_2 \text{wants-w}_0 \lf 3 \lf 4 \lf [\text{him}_s] \text{to be the next president-w}_3)]])\]

\(^{17}\) Moulton (2005) argues that not all ECM verbs support a ‘de re’/‘de se’ ambiguity of embedded subject reflexives (for example, *believe* and *consider* do not). The point made here is based on verbs that do support such ambiguity.
b. \([\text{every candidate} \ 2 \left[ t_2 \ \text{wants}^{\text{de-re-w}_0} \left[ 3 \ 4 \ [h_{im_2}] \ \text{to be the next president-w}_3 \right] \right]]\)

Clearly, (58a) does not violate the standard Condition B (i.e., the one that forbids only Type-I covaluation). And as we argued in 4.2, De Se Blocking suggests that binding by a verb cannot be ruled out by the grammar. Therefore it doesn’t matter that (58b) does violate the standard Condition B: (57) is wrongly predicted, by the standard Condition B, to be acceptable on a ‘de se’ reading of him. But according to the new Condition B, both LFs are ruled out: (58a) is ruled out because \(t_2\) and \(him_4\) are Type-II covalued. The ‘de se’ \(him\) is Type-II covalued with the trace of \(\text{every candidate}\), because it corresponds to the ‘self’ of (the denotation of) that trace.\(^{18}\)

4.4. Unexpected Condition A effects in Free Indirect Discourse

There is a contrast between first and third person reflexives in F(ree) I(ndirect) D(iscourse): only the former exhibit unexpected Condition A effects. Our proposal is compatible with this fact.

4.4.1. The contrast between first and third person in FID

FID is a form of reported speech or thought that seems to be a hybrid of S(standard) I(ndirect) D(iscourse) (embedding by an attitude verb such as say, believe, convince, suggest, dream, etc.) and quotation, as illustrated in (59) (for discussion, see Banfield 1982, Doron 1991, Schlenker 1999, 2004, Sharvit 2008, among others).

(59) “Do you love me,” asked Mary.

John answered with great passion. Yes, he loved her, and he would definitely marry her, if not today then tomorrow.

The sequence \(\text{Yes he loved her and he would definitely marry her if not today then tomorrow}\) is FID. We infer from it that John said “Yes, I love you, and I will definitely marry you, if not today then tomorrow.” If we wanted to report this using SID we would say something along the lines of \(\text{John confirmed that he loved her and said that he would definitely marry her, if not that day then the following day}\). In some respects (e.g., the fact that the speaker-oriented yes and the

\(^{18}\) More formally (see Section 3): for any relevant context \(c’\) such that \([t_2 \ \text{wants-w}_0 \left[ 4 \ 3 \ [h_{im_4} \ \text{to be the next president-w}_3] \right] \]^{c’-\phi_2_0}\) is defined and any relevant \(g’ \supseteq g_{c’}\) it holds that any \(\left< w, x \right> \in \text{Dox}(c’), \left[ t_2 \right]^{c’-\phi_2_2’}, \left\{ g’ \supseteq g’: \left[ h_{im_4} \ \text{to be the next president-w}_3 \right]^{c’-\phi_2} \text{ is defined} \right\} \cap \left\{ g’ \supseteq g’: \left[ h_{im_4} \right]^{c’-\phi_2} = x \right\} \neq \emptyset.\)
time adverbials *today* and *tomorrow* reflect John’s point of view), the FID sequence resembles the corresponding quotation. In other respects (e.g., the fact that the third person pronouns *he* and *she* take the place of *I* and *you* in the corresponding quotation), the FID sequence resembles the corresponding SID sequence.

It so happens that FID-Control constructions license third person reflexives in ‘de se’ situations but not in ‘de re’ situations. Thus, (60a) is a well-formed continuation of *Mary wished she could go back to her singing days* that uses quotation, (60b) is a well-formed continuation that uses SID, and (60c) is a well-formed FID continuation. All three continuations report a ‘de se’ thought.

(60) Mary wished she could go back to her singing days.
   a. She thought: “Ah! To hear myself sing again!”
   b. She wanted to hear herself sing again.
   c. Her mind was filled with longing. Ah! To hear herself sing again!

By contrast, the FID *To hear herself sing again* is unacceptable as reporting a ‘de re’ thought, as shown by (61c); but the corresponding SID in (61b) is acceptable (as expected, given our discussion of the SID cases involving *convince* and *promise*).

(61) Mary was listening to an old recording of hers on the radio, not recognizing her own voice.
    The radio suddenly stopped playing, and she was disappointed that couldn’t make it play again.
    a. She thought: “Ah! To hear this woman sing again!”
    b. She wanted to hear herself sing again.
    c. Her mind was filled with frustrating thoughts. *Ah! To hear herself sing again!*

    Interestingly, first person reflexives can be used to report both ‘de se’ and ‘de re’ thoughts, as illustrated in (62) and (63).

(62) A month ago, I listened to an old recording of my own voice, and wished I could go back to my singing days.
a. I thought: “Ah! To hear myself sing again!”
b. I wanted very much to hear myself sing again.
c. My mind was filled with frustrating thoughts. Ah! To hear myself sing again!

(63) A month ago, when I woke up from a coma, I didn’t know who I was. I listened to an old recording of myself, not recognizing my own voice. Suddenly, the radio stopped playing, and I was frustrated that I couldn’t make it play again.

a. I thought: “Ah! To hear this woman sing again!”
b. I wanted very much to hear myself sing again.
c. My mind was filled with frustrating thoughts. Ah! To hear myself sing again!

We now show that this contrast may be accounted for within the current proposal. As it turns out, PRO and herself in (61c) are not Type-II covalued, but PRO and myself in (63c) are.

The fact that (63c), unlike (61c), is acceptable in a ‘de re’ situation seems to be consistent with the fact that, quite generally, the reference of first person pronouns (unlike that of third person pronouns) does not shift in FID, as illustrated in (64) (see Doron 1991).19

(64) John looked at my picture, his mind filled with sad thoughts. Yes, yes, I would probably leave him soon.

John’s reported thought (while he is looking at the speaker’s picture): “This woman will probably leave me soon.”

Doron’s observation is complemented by an observation from Sharvit (2008), where it is argued that phi-features on third person pronouns in FID, unlike third person pronouns in SID, reflect the attitude holder’s perspective regarding the gender/number of the individual denoted by that pronoun; not the speaker’s. This is based on contrasts such as the contrast between the SID in (65a) and the FID in (65b), concerning the underlined pronouns.

(65) Context: John is observing Bill (who is wearing a dress and a wig) and mistakes him for a

---

19 There is some disagreement in the literature regarding the interpretation of first person pronouns in FID (specifically, between Banfield 1982 and Doron 1991). Importantly, those speakers for whom an unbound I always refers to the speaker/author of the utterance context, also accept an unexpected Condition A effect with myself.
woman. In addition, he thinks that Bill is attractive.

a. John looked at Bill. His mind was filled with exciting thoughts about the man standing in front of him, who seemed to him to be a woman. He thought that he was very attractive. He also thought he ought to ask him out.

b. John looked at Bill. His mind was filled with exciting thoughts about the man standing in from of him, who seemed to him to be a woman. Really, she/he was very attractive; he should ask her/him out.

Since (65a) is an instance of SID – embedding under an attitude verb (think) – he is appropriate: it is a ‘de re’ pronoun which reflects the speaker’s beliefs regarding the gender of the referent of he. On the other hand, (65b) is an instance of FID; here, she is the appropriate pronoun to use because it reflects the beliefs of John (who mistakes Bill for a woman).

Following a suggestion by Doron (1991) (later developed in Sharvit 2008), let us assume that FID involves quantification over contexts and assignments compatible with what the attitude holder believes, so that the embedded clause is interpreted relative to those contexts and assignments. Attitude verbs, by contrast (and at least in English) quantify over world-individual pairs compatible with what the attitude holder believes (and crucially not over assignments). For simplicity, let us assume (following an idea from Schlenker 1999) that this quantification is done via an FID-operator. This operator is a special context-shifter, and among the items whose reference it shifts are free third person pronouns. When the context is shifted, a third person pronoun that is free in the complement clause of the FID-operator receives a reference that reflects what the attitude holder believes (not what the speaker believes), regarding the gender of the referent. But in order to capture Doron’s observation regarding first person pronouns, let us also assume the following: (i) every node is interpreted relative to two context parameters – c and Q – where Q is always the utterance context, and in the default case (i.e., when no context-shifting takes place), c = Q (as implied by (66a)); and (ii) free first person pronouns always refer

---

20 A minor technical clarification is in order: we have been following the practice of calling a triple consisting of a world, a speaker/author and an assignment – a context. Often (e.g., Schlenker 1999), pairs consisting of a world and a speaker/author are called contexts. In the latter sense of ‘context’, attitude verbs certainly quantify over contexts. But the main distinction in the theory advocated in Sharvit (2008) is between operators that quantify over variable assignments and operators that do not. English attitude verbs do not quantify over variable assignments (so the pronominal elements that are free in the embedded clause receive the same value they would if they appeared “upstairs”). By contrast, the FID operator quantifies over assignments (so the pronominal elements that are free in the embedded clause may receive an interpretation that is different from the one they would receive “upstairs”).
to speaker/author(Q) (as implied by (66b)). So although we know that Bill is a man, the right
pronoun to use in (65b) is *she*, which carries the presupposition – reflecting John’s beliefs – that
the referent is a female, as shown in (67), which relies on the semantics of *FID* given in (66c).
This semantics is based on the assumption – introduced in Section 2.1 – that for any context c,
the range of $g_c$ is a subset of those semantic objects that are salient for author/speaker(c) and
reflects the beliefs of author/speaker(c)) regarding, for example, the gender of certain individuals
(where K is the set of possible contexts, $[[\alpha]]^{Q,g} = [\lambda c' \in K . [[\alpha]]^{c',Q,g_c}];$ adapted from

(66) a. If $\alpha$ is a root node, then $[[\alpha]]^{Q,g}$ is defined only if $c = Q$ and $g = g_c$.

b. $[[I_k]]^{Q,g}$ is defined only if $g(k) = \text{author/speaker}(Q)$.

21 The phi-features of $I$ may sometimes (e.g., *Only I got a question I understood*, see Heim 2005 and Kratzer 2009)
be invisible to semantic interpretation (in the sense of Footnote 4), in which case $I$ does not denote speaker(Q).

c. For any function $f \in D_{<k,<<e,t>>}$ (from K to $D_{<s,<<e,t>>}$), any $x \in D_e$ and any $w \in D_s$,
$[[FID]]^{Q,g}(w)(x)(f)$ is defined only if for all $c' \in K$ (i.e., for all $<w', x', g'> \in K$ such
that $w' = \text{world}(c')$, $x' = \text{author/speaker}(c')$ and $g'$ is the assignment supplied by $c'$)
compatible with what x believes in $w$, $f(c')(\text{world}(c'))(\text{speaker/author}(c'))$ is defined.
Whenever defined, $[[FID]]^{Q,g}(w)(x)(f) = True$ iff for all $c'$ compatible with what x
believes in $w$, $f(c')(\text{world}(c'))(\text{speaker/author}(c')) = True$.

(67) a. $[FID-w_0-John^2][3 1 [L_D she_2 be-w_3 attractive]]$

b. Whenever defined (i.e., when for all $c'$ compatible with what John believes in
world(Q), $[[\text{she}_2 be-w_3 attractive}]^{c',Q,g_c[3\rightarrow\text{world}(c'), 1\rightarrow\text{speaker/author}(c')]}$ is defined; which means
that at the very least, $g_c(2)$ is a female in world(c')),
$[[67a]]^{c',Q,g_c} = True$ iff for all $c'$ compatible with what John believes in world(Q): $[[\text{she}_2 be-w_3 attractive}]^{c',Q,g_c[3\rightarrow\text{world}(c'), 1\rightarrow\text{speaker/author}(c')]} = True$ (i.e., for all $c'$ compatible with
what John believes in world(Q), $g_c(2)$ is attractive in world(c')).

Now, going back to the contrast between first and third person pronouns, recall that in a ‘de
se’ context where Mary is listening to her voice and recognizes it, both (68a) and (68b) are
acceptable. But in the ‘de re’ context in (61), where Mary doesn’t realize that she is the singer, (68a) is acceptable but (68b) is not. (68c), on the other hand, with a first person pronoun, is a acceptable in both ‘de se’ and ‘de re’ situations.

(68) a. She wanted to hear herself sing again.
b. Ah! To hear herself sing again.
c. Ah! To hear myself sing again.

Since pronouns that are free in the clausal complement of $FID$ are interpreted relative to the context (or contexts) and assignment (or assignments) introduced by $FID$, a reflexive pronoun that is free in the clausal complement of $FID$ may have a semantic value that is different from that of the NP that syntactically binds it, even when they are co-indexed. This forces us to slightly amend the definition of Type-I covaluation: taking into account that the syntactic tree may contain a context-shifter, and taking into account that different NPs may therefore be interpreted relative to different assignments, Type-I covaluation is now defined as in (69) and relies on the term ‘adequate’ (defined recursively in (70)). The idea is that an assignment is adequate for an NP c-commanded by $FID$ if it provided by a context quantified over by $FID$.

(69) For any root LF $\alpha$ and context $c$ such that $\llbracket \alpha \rrbracket^{c,Q,g}$ is defined: any two NPs (of type e), $\beta$ and $\gamma$, that are dominated by $\alpha$ are Type-I covalued in $\alpha$ relative to $c$ iff:

(i) there is at least one assignment $g$ adequate for both $\beta$ and $\gamma$ in $c$ relative to $\alpha$ and $Q$;
(ii) for all assignments $g$ adequate for $\beta$ and $\gamma$ in $c$ relative to $\alpha$ and $Q$, for any $g' \supseteq g$ such that $\llbracket \gamma \rrbracket^{c,Q,g'}$ and $\llbracket \beta \rrbracket^{c,Q,g'}$ are defined, $\llbracket \gamma \rrbracket^{c,Q,g'} = \llbracket \beta \rrbracket^{c,Q,g'}$.

(70) Assignment $g$ is adequate for $\alpha$ in $c$ relative to node $\beta$ and $Q$ iff (i) or (ii) holds:

(i) $\alpha$ is not c-commanded by any context-shifters in $\beta$, and $g = g_c$;

(ii) $\alpha$ is c-commanded by at least one context-shifter in $\beta$, and the closest context-shifter c-commanding $\alpha$, $S^w$-w$^a$-NP$^w$ (where $S^w$ could stand for $FID$, w$^a$ is the world-argument of $S^w$ and NP$^w$ its attitude-holder-denoting argument), is such that: (a) $\alpha$ is dominated by the clausal complement of $S^w$-w$^a$-NP$^w$; and (b) $g$ is compatible, as far as $\alpha$ is concerned, with what $\llbracket NP^w \rrbracket^{c,Q,g''}$ believes in $\llbracket w^a \rrbracket^{c,Q,g''}$ (where $g'' \supseteq g'$ and $g'$ is an assignment.
adequate for $S^\alpha$-w$^\alpha$-NP$^\alpha$ in c relative to $\beta$ and Q).

When both NPs are c-commanded by the same context-shifters (and this includes the case where no context-shifter c-commands either of them), Type-I covaluation is determined as before and always holds when the two NPs are co-indexed pronouns. When there is at least one context-shifter intervening between the two NPs, Type-I covaluation need not hold even if the two NPs are co-indexed pronouns. This is because it is possible that there are no assignments that are adequate for both of them.

Let us also adjust the (simplified version of the) definition of Type-II covaluation in (24), to cover FID cases as well.

(71) For any root node $\alpha$ and any attitude verb/FID-operator ATT such that $\alpha$ dominates ATT, let: (i) $w(ATT)$ be the pronominal world-argument of ATT, (ii) $S(ATT)$ be the first node that dominates ATT and all its syntactic arguments, (iii) $R(ATT)$ be the highest node in the clausal complement of ATT that is c-commanded by both abstraction indices introduced by ATT, and (iv) $k(ATT)$ be the world-binding abstraction index introduced by ATT and $m(ATT)$ be the individual-binding attraction index introduced by ATT.

Then –

Any two NPs (of type e), $\beta$ and $\gamma$, that are dominated by root $\alpha$ are Type-II covalued in $\alpha$ if and only if there is at least one ordered pair $<$NP1, NP2$>$ $\in$ {$<$ $\beta$, $\gamma$, $\gamma$, $\beta$ $>$} such that for any $c_1$ such that $g_{c_1}$ is appropriate for $S(ATT)$ and $[$$S(ATT)$$]$[$^1.Q,g_{c_1}$] is defined, $g_{c_1}$ is adequate for NP1 in $c_1$ relative to $S(ATT)$ and Q and:

(i) if ATT is an attitude verb, then for any $g' \supseteq g_{c_1}$ such that $\text{Dom}(g') = \text{Dom}(g_{c_1}) \cup \{j: j$ is a pronominal index that is free in NP1} and $[$$NP1$$]$[$^1.Q,g'$] is defined, for any $<w$, x> compatible with what $[$$NP1$$]$[$^1.Q,g'$] believes in $[$$w(ATT)$$]$[$^1.Q,g'$]:

$\{g'' \supseteq g': [R(ATT)] [^1.Q,g'']$ is defined$\} \cap \{g'' \supseteq g'[k(ATT) \rightarrow w]: [NP2] [^1.Q,g''] = x \} \neq \emptyset$; and

(ii) if ATT is FID, then for any $g' \supseteq g_{c_1}$ such that $[$$NP1$$]$[$^1.Q,g'$] is defined, for any $c_2$ compatible with what $[$$NP1$$]$[$^1.Q,g'$] believes in $[$$w(ATT)$$]$[$^1.Q,g'$] (such that $g_{c_2}$ is appropriate for the complement clause of FID):

$\{g'' \supseteq g_{c_2}: [R(ATT)] [^2.Q,g'']$ is defined$\} \cap \{g'' \supseteq g_{c_2}[k(ATT) \rightarrow \text{world}(c_2)]: [NP2] [^2.Q,g''] =}$
When ATT is an attitude verb, Type-II covaluation is determined as before (see Section 3). When ATT is the FID-operator, Type-II covaluation can never hold between PRO and a free third person pronoun in the clausal complement of ATT, because Type-II covaluation requires the free NP to denote the attitude holder in all relevant contexts $c_1$ such that $[[S(ATT)]]^{c_1,c,g}$ is defined. Let us see why this cannot be, assuming that the LF of To hear herself sing again! is (72).

(72) *[FID-$w_0$-Mary/she$_6$ $\wedge$[2 1 [LD [NP$_2$ PRO$_1$] hear-$w_2$ [NP$_1$ herself$_6$] sing again]]]

Only $g_c$ is adequate for she$_6$ in Q. Since Mary is listening to herself on the radio (in the situation described in (61)), $g_c(6) =$ Mary. In the situation described, Mary thinks someone else is the singer. This means that $g_c$ is not adequate for herself$_6$, because it is not compatible with what $g_c(6)$ (i.e., Mary) believes in world(c). This means that herself$_6$ and she$_6$ are not covalued; neither are PRO and herself$_6$.

Importantly, even if for some utterance context, herself$_6$ and she$_6$ are Type-I covalued, (72) is still ruled out. Here is why. Imagine a context, call it $c'$, such that herself$_6$ and Mary/she$_6$ are Type-I covalued relative to $c'$. At the very least, $c'$ must be such that Mary recognizes the singer as Mary (although she may not think that she herself is the singer, because she may not know who she is). Type-II covaluation still cannot hold between PRO$_1$ and herself$_6$, because (71) requires herself to denote the attitude holder in all relevant contexts $c_1$ such that $[[S(ATT)]]^{c_1,c,g_c}$ is defined. We already know that at least one such context (namely, the one where Mary thinks that the singer is Sally) is not like that.

Now, consider Ah! To hear myself sing again!, which is well-formed in both ‘de re’ and ‘de se’ situations. Its ‘de re’ LF is (73).

(73) [FID-$w_0$I$_6$ $\wedge$[2 1 [LD [NP$_2$ PRO$_1$] hear-$w_2$ [NP$_1$ myself$_6$] sing again]]]

Because myself$_6$, by stipulation (see (66b)), denotes speaker/author(c) (where $c$ is the utterance context), in any context, it follows that for any context $c'$ compatible with what $g_c(6)$ thinks in
world(c), g_c(6) = g_c(6) = speaker/author(c). Therefore, by (71), PRO and myself are Type-II covalued.

4.4.2. Third person pronouns and the utterance context

We have just seen that the contrast between first and third person reflexives in FID is predicted by the theory that accounts for the fact that phi-features on a third person pronoun in FID reflect the attitude holder’s point of view. It is important to guarantee that our theory also accounts for the fact that in a different respect (i.e., not related to phi-features), third person pronouns reflect the speaker’s point of view. This is illustrated by (74)-(75) (where the situations are ‘de re’ situations) which contrast with (76)-(77) (where the situations are ‘de dicto’ situations).

(74) John was staring at the dean, who is a woman. He didn’t recognize her and mistakenly thought he was looking at the provost (who is actually a man, but John thinks he is a woman). He was convinced, while pointing at the dean, that she loved him.
   a. He was obviously wrong; the dean didn’t love him at all (though the provost did).
   b. #He was obviously wrong; the provost didn’t love him at all (though the dean did).

(75) John was staring at the dean, who is a woman. He didn’t recognize her and mistakenly thought he was looking at the provost (who is actually a man, but John thinks he is a woman). Yes, yes, she loved him, he kept telling himself as he pointed at the dean.
   a. He was obviously wrong; the dean didn’t love him at all (though the provost did).
   b. #He was obviously wrong; the provost didn’t love him at all (though the dean did).

(76) John doesn’t know who the provost – who is a man – is. He is convinced that the provost/(#she) loves him. He is obviously wrong; the provost doesn’t love him at all.

(77) John doesn’t know who the provost – who is a man – is. Yes, yes, the provost/she loves him, he keeps telling himself. He is obviously wrong; the provost doesn’t love him at all.

It seems that John “is wrong” in the SID in (74) and the FID in (75) about roughly the same thing; likewise, John “is wrong” in the SID in (76) and the FID in (77) about roughly the same thing. To account for the fact that FID and SID do not differ from each other in this particular
respect, we should recognize that just like *convince* has a ‘de re’ variant (see (23)), so does the FID-operator, whose ‘de re’ variant is given in (78) (see also Sharvit 2008).\footnote{22,23}

(78) For any \( \langle h, f \rangle \) such that \( h \in D_{k,e} \) (the set of functions from \( K \) to \( D_e \)) and \( f \in D_{k,e,\langle\langle\rangle\rangle} \), \([FID]^{de-re}_{c,q,g}(x)(w)(\langle h, f \rangle)\) is defined only if \( c \) supplies a function in \( D_{k,e} \), \( m_c \), such that \( m_c \supseteq h \), and a suitable description in \( D_{e,\langle\langle\rangle\rangle} \), \( F^h_c \), such that:

a. for any \( c' \in K \) compatible with what \( x \) believes in \( w \):
   
   i. \( F^h_c(\text{author}(c'))(\text{world}(c')) = h(c') \);
   
   ii. if \( m_c(c) \) is defined, then \( F^h_c(x)(w) = m_c(c) \); and
   
   iii. (a) if \( h(c') = \text{author}(c') \) then \( m_c(c) = x \), and
   
   (b) if \( h(c') = \text{world}(c') \) then \( m_c(c) = w \).

b. \( f(c')(F^h_c(\text{author}(c'))(\text{world}(c')))(\text{world}(c'))(\text{author}(c')) \) is defined.

Whenever defined, \([FID]^{de-re}_{c,q,g}(x)(w)(\langle h, f \rangle) = True \) iff for all \( c' \) compatible with what \( x \) believes in \( w \), \( f(c')(F^h_c(\text{author}(c'))(\text{world}(c')))(\text{world}(c'))(\text{author}(c')) = True \).

Accordingly, the clausal internal argument of the FID-operator is not the denotation of its clausal complement, but rather a pair – called \( \langle h, f \rangle \) in (78) – where \( h \) is formed by taking a “free” individual-denoting expression in the clausal complement and constructing the appropriate function in \( D_{k,e} \) (taking into account the presuppositions imposed by the phi-features of that expression, if any). Thus, when (67a) (i.e., \([FID]^{de-re}_{w_0 John \ n[3 1 [\ldots she \ be-w, attractive ]]]} \)) is evaluated relative to context \( c \) (= \( Q \); the context described in (65)), then \( h = [\lambda c'' \in K . g_{c'}(2) \text{ is a female in world}(c'')] \). \([she_2]\)\(^{c,q,e_c}\), and the clausal internal argument of \( FID^{de-re} \) is the pair \( \langle [\lambda c'' \in K . g_{c'}(2) \text{ is a female in world}(c'')] \rangle. m_c \) has to be some superset of \( [\lambda c'' \in K . g_{c'}(2) \text{ is a female in world}(c'')] \). \([she_2]\)\(^{c,q,e_c}\), so suppose \( m_c = [\lambda c'' \in K . g_{c'}(2)] \). For \( [([67a])]^{c,q,e_c} \) to be defined, it must be the

\footnote{22 (78) may require us to treat \( I \) as assignment-independent (i.e., a “real” indexical): \( [I]^{Q,e} = \text{author}(Q) \).}

\footnote{23 Philippe Schlenker points out that without an overt parenthetical (e.g., *he kept telling himself*), an FID cannot be continued with *He was wrong* (e.g., *She loved him. He was obviously wrong*). While this might mean that we have to refine the semantics of *FID* (combining insights from Schlenker 2004), it doesn’t undermine the claim that the semantics of *FID* has a ‘de re’ variant.}
case that: (i) if $[\lambda c'' \in K . g_c(2)](c)$ is defined, $[\lambda c'' \in K . g_c(2)](c) = F_c^h(\text{John})(\text{world}(c))$, and (ii) for all $c'$ compatible with what John believes in $\text{world}(c)$, $[\lambda c'' \in K . g_c''(2)](c) = F_c^b(\text{author}(c'))(\text{world}(c'))$. In other words, (78) guarantees that 2 – the free index of she – is visible to both “matrix” and “embedded” assignments. More generally, the index of any free pronoun, not only a first person pronoun, is in principle visible to both “matrix” and “embedded” assignments. But the phi-features of a third person pronoun are still visible only to the “embedded” assignments. Accordingly, we get (79)-(82) for (74)-(77) (assuming $c = Q$, the utterance context).

(79) $[\text{John be-convinced}^{de-re-w_0}[35[\text{she}_7\text{love-w}_3\text{him}_3]]]$  
(74)

By (23), this LF presupposes a description $F_c^h$ (e.g., $[\lambda x \in D_x . \lambda w \in D_s . \text{[love]}(w)(\text{John})(y)]$) such that $g_c(7) = F_c^h(\text{John})(\text{world}(c))$. John is in fact pointing at the dean.

(80) $[\text{FID}^{de-re-w_0-John}^*[35[\text{she}_7\text{love-w}_3\text{him}_3]]]$  
(75)

By (78), this LF presupposes a description $F_c^h$ such that $g_c(7) = F_c^h(\text{John})(\text{world}(c))$ and for all context $c'$ compatible with what John believes in $\text{world}(c)$, $g_{c'}(7) = F_c^b(\text{author}(c'))(\text{world}(c'))$. John is in fact pointing at the dean.

(81) $[\text{John be-convinced-w}_0[35[\text{the provost-w}_3\text{love-w}_3\text{him}_3]]]$  
(76)

(82) $[\text{FID}^{de-re-w_0-John}^*[35[\text{she}_7\text{love-w}_3\text{him}_3]]]$  
(77)

Now we can account for the fact that what John is wrong about in (74) and in (75) corresponds to the structured proposition in (83a) (where $g_c(7)$ is the dean in $\text{world}(c)$), and what he is wrong about in (76) and (77) is (83b).

(83) a.  $<g_c(7), [\lambda y \in D_y . \lambda w \in D_s . [[\text{love}]](w)(\text{John})(y)>$

b.  $[\lambda w \in D_s . [[\text{love}]](w)(\text{John})(\text{the provost in w})]$

(83a) can only be obtained for (80) if $g_c$ has access to the indices in the complement clause, as we now assume. Thus, our theory correctly captures the fact that free pronouns are “visible” to the assignment supplied by the utterance context, even in complement clauses of FID.

Notice that we predict the SID in (84a) and the FID in (84b) (where the underlined she refers to Mary), to imply that Mary “is wrong” about the same thing (namely, (84c)). Likewise,
in the SID in (85a) and the FID in (85b) (where the underlined she refers to Mary and the boldface she to her ‘self’) to imply that Mary “is wrong” about the same thing (namely, (85c)). The same English sentence – *Mary looks ugly* – expresses what Mary is wrong about.

(84) Mary looked at a picture of herself without recognizing herself in the picture.
   a. **She** was convinced **she** looked ugly. **She** was obviously wrong; *Mary didn’t look ugly at all.*
   b. **Really! She** looked ugly, **she** kept telling herself. **She** was obviously wrong; *Mary didn’t look ugly at all.*
   c. <g_she> (=Mary) , [λy ∈ D_y . λw ∈ D_w ∧ x ∈ D_x . ![look-ugly](w)(y)]>

(85) Mary looked at a picture of herself, recognizing herself.
   a. **She** was convinced **she** was looked ugly. **She** was obviously wrong; *Mary didn’t look ugly at all.*
   b. **Really! She** looked ugly, **she** kept telling herself. **She** was obviously wrong; *Mary didn’t look ugly at all.*
   c. ![look-ugly](w)(x)>

This does not undermine our conclusion from Section 4.4.1: the values that the “matrix” and “embedded” assignments actually assign to the same pronoun may be different, and as a result, third person ‘de re’ reflexives are still predicted to be excluded from complement clauses of FID (see (72)) because, as we saw, they always fail to satisfy Condition A (i.e., they never come out as either Type-I or Type-II covalued with the embedded subject).

### 5. Alternative theories of unexpected BT effects

In this section we explore three alternative theories of unexpected BT effects and the challenges they face. What these alternatives have in common is that they account for our basic example (*McCain convinced Palin to vote for herself*) without complicating BT with reference to Type-II covaluation. However, while some of them account for some non-basic facts (discussed in Section 4), they all seem to fail to account for the entire range of facts discussed here. It is not our purpose here to claim that alternative theories of unexpected BT effects are not possible, but rather to make clear what challenges such theories have to face.
5.1. The pure ‘de re’ theory

In this section we ask whether the pure ‘de re’ LF in (86a), where $\text{PRO}$ is co-indexed with $\text{herself}$ (satisfying the standard Condition A) but not bound by $\text{convince}$, is a well-formed source of the reading of \textit{McCain convinced Palin to vote for herself} where $\text{herself}$ is interpreted ‘de re’ and (86b) (which violates the standard Condition B), the only source of the reading of \textit{McCain convinced Palin to vote for her}, where \textit{her} is interpreted ‘de re’.

(86)  
\begin{align*}
\text{a. } & \text{[Palin } 6 \text{ [McCain convinced}\text{de-re-}w_0 t_6 [3 5 [\text{LD } \text{PRO}_6 \text{ vote for-}w_3 \text{ herself}_6]\text{]]]}
\text{b. } & \ast\text{[Palin } 6 \text{ [McCain convinced}\text{de-re-}w_0 t_6 [3 5 [\text{LD } \text{PRO}_6 \text{ vote for-}w_3 \text{ her}_6]\text{]]]}
\end{align*}

Likewise, we ask whether (87a) is a well-formed source of the reading of \textit{McCain convinced Palin that she had voted for herself}, where \textit{she} is understood ‘de se’ and \textit{herself} ‘de re’, and (87b) the only source of the reading of \textit{McCain convinced Palin that she had voted for her}, where \textit{she} is understood ‘de se’ and \textit{her} ‘de re’.

(87)  
\begin{align*}
\text{a. } & \text{[Palin } 6 \text{ [McCain convinced}\text{de-re-}w_0 t_6 [3 5 [\text{LD } \text{she}_6 \text{ vote for-}w_3 \text{ herself}_6]\text{]]]}
\text{b. } & \ast\text{[Palin } 6 \text{ [McCain convinced}\text{de-re-}w_0 t_6 [3 5 [\text{LD } \text{she}_6 \text{ vote for-}w_3 \text{ her}_6]\text{]]]}
\end{align*}

The answer is that this pure ‘de re’ approach is not general enough. The arguments come both from SID (where the pure ‘de re’ approach doesn’t cover Condition B effects) and FID (where it doesn’t account for the fact that first person reflexives have ‘de re’ readings). Let us start with the former.

Recall from Section 4.1 that $\text{PRO}$ is an inherently ‘de se’ pronoun. Thus, for (86a) to count as a possible LF, we have to stipulate that when $\text{PRO}$ is a ‘de re’ pronoun embedded under an attitude verb, it can only be interpreted relative to the identity function (otherwise we wouldn’t predict it to be an inherently ‘de se’ pronoun), while other pronouns may be interpreted relative to other descriptions. In order to guarantee that (86b) is the only possible LF for \textit{McCain convinced Palin to vote for her} (and that a Condition B violation cannot be avoided by an LF where $\text{PRO}$ is bound by $\text{convince}$ and $\text{her}$ is free), we have to stipulate further that $\text{PRO}$ is always syntactically ‘de re’ (i.e., never bound by an attitude verb or the FID operator).
Furthermore, to claim the same for (87b) we have to stipulate that ALL ‘de se’ pronouns, not only \textit{PRO}, are syntactically ‘de re’ pronouns interpreted relative to the identity function (this view is advocated in Lewis (1979) and Reinhart (1991), among others).

One reason to reject the pure ‘de re’ approach to unexpected BT effects is that taking into account De Re and De Se Blocking, unexpected Condition A effects may be predicted, but it is hard to see how unexpected Condition B effects can be predicted. As we saw in Section 4.3, De Se Blocking led us to the conclusion that the grammar cannot block binding of pronouns by a verb. This, in turn, means that we have to acknowledge that while \textit{McCain convinced Palin to vote for her} or \textit{McCain convinced Palin that she had voted for her} (where \textit{Palin} and \textit{her} are understood as Type-I covalued) certainly have LFs where the standard Condition B is violated, they also have LFs where the standard Condition B is not violated (i.e., LFs where \textit{PRO/she} are bound by \textit{convince} and \textit{her} is not). For the very same reason, unexpected Condition B effects with ‘de se’ pronouns (Sections 4.1 and 4.3) cannot be accounted for either.

Another reason to reject the pure ‘de re’ approach concerns the behavior of first person pronouns in FID, which also shows that syntactically ‘de se’ pronouns are indeed generated by the grammar. As we saw in section 4.4.1, first person reflexives (unlike third person reflexives) induce unexpected Condition A effects in FID. The problem is that just like we have to stipulate for SID that \textit{PRO} must be interpreted relative to the identity function, for FID we have to stipulate that \textit{PRO} must denote, in any context quantified over by \textit{FID}, the ‘self’ of the attitude holder. This is because \textit{FID} is a context-shifter. This, in turn, means that in (88), \textit{myself} – which is co-indexed with \textit{PRO} – must also refer to the ‘self’ of the attitude holder.

(88) \[
[FID^{de-re-w_0-I} \ ^\wedge [3 \{_{LD} PRO_6 \text{hear-}w_3 \text{myself}_6 \text{sing again}\}]]
\]

In other words, (88) is, in principle, incompatible with a ‘de re’ situation. We are therefore forced to conclude that in order to account for the inherent ‘de se’-ness of \textit{PRO}, and at the same time account for the fact that first person pronouns in FID give rise to unexpected Condition A effects, we have to assume that \textit{PRO} is bound by the FID-operator, as in (89).

(89) \[
[FID^{de-re-w_0-I_6} \ ^\wedge [3 \{_{LD} PRO_1 \text{hear-}w_3 \text{myself}_6 \text{sing again}\}]]
\]
And since PRO can be bound by FID, there is no obvious theoretical reason to say that it cannot be bound by an attitude verb in SID. This conclusion is compatible with our conclusion regarding De Se Blocking: we can easily account for it assuming that an embedded subject pronoun – overt or covert – may be locally bound.

5.2. The subject-deletion theory

An explanation that represents this approach is sketched in Heim (1994a) (see also Sauerland 2001). The idea behind this approach is that a subject ‘de se’ pronoun may optionally be deleted (along with the individual-binding abstraction index that binds it), when c-commanded by a morphologically agreeing expression. What makes this deletion possible is that fact that [1 [pronoun, [vp α]]] and [vp α] have identical denotations. For example, [[3 5 [PRO, vote for-w3 herself6]]] $\uparrow^{\text{C}(\text{e-ry})} = [[3 [vote for-w3 herself6]]]$, so PRO and 5 can be deleted from [3 5 [PRO, vote for-w3 herself6]]. At the same time, getting rid of a piece of the tree enables the relevant local domain to include both the reflexive pronoun and the relevant trace.

(90) a. No PRO-deletion: (The standard) Condition A is violated unless PRO and herself are co-indexed. Result: ‘de se’ reading of herself.

[McCain convinced-w0 Palin [3 5 [LD PRO, vote for-w3 herself6]]]

b. PRO-deletion: (The standard) Condition A is violated unless the trace of Palin and herself are co-indexed. Result: ‘de re’ reading of herself.

[Palin [6 [McCain convinced-re-w0 [LD t6 [3 [vote for-w3 herself6]]]]]]

There are several problems with this approach, of which we address only two (see Charlow 2009 for additional criticism). Starting with Condition B effects with object ‘de re’ pronouns, let us consider McCain convinced every female politician to vote for her, on the assumption that PRO-deletion is optional. We predict (91a), where PRO-deletion has taken place, to be ungrammatical, but (91b), where PRO-deletion has not taken place, to be grammatical. We should, of course, predict them both to be ungrammatical (see (13)).

(91) a. *[every politician [2 [McCain convinced-re-w0 [LD t2 [3 [vote for-w3 her2]]]]]]

b. [every politician [2 [McCain convinced-re-w0 t2 [5 3 [LD PRO, vote for-w3 her2]]]]]
We could, of course, stipulate that PRO-deletion is obligatory. Optional or obligatory, PRO-deletion renders (92) – with a ‘de re’ reflexive pronoun, which doesn’t violate the standard Condition A – a well-formed LF of the unacceptable McCain promised Palin to vote for herself.

(92) \[ \text{Palin [2 [McCain promised^{de-re}-w_0 t_2 [3 [vote for-w_3 herself]]]]] } \]

This LF gives rise to the reading where McCain promises Palin, while pointing at a picture of her: “I will vote for you/this woman.”

The subject-deletion theory also fails to capture unexpected Condition B effects with object ‘de se’ pronouns. McCain convinced every female politician that she had voted for her is predicted to be well-formed (on the reading where she is not bound by convince and her is ‘de se’; see (39)): the object pronoun cannot be deleted (because [5 3 [she vote for-w_3 her_3]] and [3 [she vote for-w_3]] do not have the same denotation). Hence, both she and her are free in their local domain, in satisfaction of the standard Condition B.

(93) \[ \text{every politician [2 [McCain convinced^{de-re}-w_0 t_5 [5 3 [LD she vote for-w_3 her_3]]]] } \]

Condition B effects with subject ‘de se’ pronouns in ECM constructions (discussed in 4.3) cannot be accounted for either. For example, Biden wants him to get re-elected is predicted to be ruled in on a ‘de se’ reading of him.

5.3. The extensional BT theory

It is suggested in Charlow (2010a,b) that BT “cares” only about extensional values of NPs (that is to say, the values that NPs have when evaluated outside the scope of any intensional operators). Let us briefly discuss the challenges for this idea (without being entirely faithful to what is actually proposed in Charlow 2010a,b).

Suppose we somehow make it a requirement (perhaps in the form of a presupposition) that the extensional value of PRO be the same as that of the c-commanding attitude-holder-denoting NP. Regarding (94), this means that for any x in the domain of \( [\lambda y \in D_e. [[McCain convinced^{de-re}-w_0 t_6 [3 5 [LD \text{PRO}_5 \text{ vote for-w}_3 \text{ herself}_6]]]]] \), \( [\text{PRO}_5]^{\text{de-int}[5 \to x]} = [t_6]^{\text{de-int}[6 \to x]} \). Taken together
with the assumption that Conditions A and B care only about extensional values, we predict that Condition A is satisfied in (94a) and Condition B is violated in (94b).

(94) a. \([\text{Palin [6 [McCain convinced}^{de-re-w_0}t_6 [3 5 \text{ [}_{{\text{LD}} \text{PRO}_3} \text{ vote for-w_3 herself}_6]]]]]\)
   b. \([\text{Palin [6 [McCain convinced}^{de-re-w_0}t_6 [3 5 \text{ [}_{{\text{LD}} \text{PRO}_3} \text{ vote for-w_3 her}_6]]]]]\)

Furthermore, we easily predict ‘de se’ non-reflexive pronouns to respect Condition B, as illustrated in (95): as long as we can force the extensional value of \(\text{him}\) to be that of the relevant attitude holder, it will have the same extensional value as \(t_2\).

(95) \([\text{every candidate [2 [}_{{\text{LD}} t_2 \text{ wants-w_0 [3 4 [\text{him}_4] to be the next president-w_3]]]]]}\]"

The main challenge for this theory is the contrast between first and third person pronouns in FID. Recall (68b) (Ah! To hear herself sing again!), which is unacceptable in the ‘de re’ situation described in (61), and its first person counterpart in (68c) (Ah! To hear myself sing again!) which is acceptable in the corresponding ‘de re’ situation. Assume the following LFs.

(96) a. \([\text{FID}^{de-re-w_0-Mary} [3 1 [}_{{\text{LD}} \text{PRO}_1 \text{ hear-w_3 herself}_6 \text{ sing again}]]]\]
   b. \([\text{FID}^{de-re-w_0-I_6} [3 1 [}_{{\text{LD}} \text{PRO}_1 \text{ hear-w_3 myself}_6 \text{ sing again}]]]\]

The extensional values of \(\text{herself}\) and \(\text{myself}\) are determined by the assignment of the utterance context, not the contexts quantified over the FID-operator. On the assumption that the extensional value of \(\text{PRO}\) equals that of \(\text{Mary}\), we get the same results as in (94a), for both \(\text{herself}\) and \(\text{myself}\). Any attempt to amend the theory in order to derive a contrast between \(\text{herself}\) and \(\text{myself}\) would also have to take into account the fact that phi-features on third person pronouns are interpreted from the attitude holder’s point of view (see Section 4.4.1), but at the same time, the indices of third person pronouns and first person pronouns alike are visible to the matrix assignment (see Section 4.4.2).

To conclude, all the alternatives explored in this section have the advantage that they do not resort to Type-II covaluation, yet they face some challenges. The subject-deletion theory fails to capture Condition B effects, as does the pure ‘de re’ theory, which also fails to capture the fact
that free first person reflexives are allowed in FID. The extensional BT theory has to deal with
special properties of third person pronouns in FID. The theory that says that BT is sensitive to
both Type-I and Type-II covaluation accounts for all these facts.

7. Endnote
Let us end with a puzzle. Consider the expected and unexpected Condition C effects in (97).

(97) Palin believes that McCain voted for the former female governor of Alaska.

The standard Condition C disallows Type-I covaluation between a name or a definite description
and a c-commanding NP. In a context such as ours, where Palin and the former female governor-
w_0 of Alaska are Type-I covalued, the standard Condition C correctly rules (97) out on a ‘de re’
reading of the former female governor of Alaska. On the ‘de dicto’ reading of the former female
governor of Alaska, the acceptability of (97) depends on Palin’s state of mind. If she has no idea
who the former female governor of Alaska is, the sentence is acceptable. But if she knows that
she is the former female governor of Alaska, the sentence is odd. In some such cases where (97)
is odd, Palin may not know who she is (while still believing that she is the former female
governor of Alaska). That kind of oddity is an unexpected Condition C effect because when
Palin doesn’t know who she is, the former female governor of Alaska need not pick out the same
individual in all of her doxastic alternatives. In that case, Palin and the former female governor
of Alaska are not Type-I covalued (yet the sentence is not fully acceptable). Interestingly, Palin
and the former female governor of Alaska are not Type-II covalued either, because there are
bound to be contexts c’ where she doesn’t think she is the governor of Alaska.

It is conceivable that the oddity of (97) in the second ‘de dicto’ situation described above is
independently accounted for by a pragmatic preference not to self-ascribe (or report self-
ascription by others) using a name or a definite description (for example, Sarah Palin probably
does not say The former female governor of Alaska is hungry or even Sarah is hungry when she
wants to convey the message that she is hungry). This, in turn, means that some unexpected BT
effects (specifically, Condition C effects) are explained without reference to Type-II covaluation.

Appendix
The Type-II approach to unexpected BT effects presented in Section 3 must be refined a bit in order to account for additional facts, especially facts concerning a ‘de te’, rather than ‘de se’, interpretation of PRO. This is illustrated by Control verbs that support such readings (we ignore FID cases, for simplicity).

For example, *McCain told Palin to vote for herself* should be ruled in, on a ‘de re’ reading of *herself*. Likewise, *McCain talked to Palin about advertising themselves* should be ruled in on a ‘de re’ reading of *themselves* (the latter observation is documented in Charlow 2009). *Tell*, like *convince*, is Object Control, but it differs from *convince* in that it quantifies over different doxastic alternatives. While *converse* quantifies over pairs <w, x>, where x is an individual the ‘object’ considers a plausible candidate for who she/he is, *tell* quantifies over triples <w, x, y>, where x is ‘self’ – an individual that the ‘subject’ considers a plausible candidate for who he/she is, and y is ‘addressee-self’ – an individual that the ‘subject’ considers a plausible candidate for who he/she is addressing. *Tell* supports a ‘de re’ / ‘de te’ ambiguity, as shown in (1) (unlike *promise*, which supports a ‘de re’ / ‘de se’ ambiguity), and *talk-to* supports a ‘de re’ / ‘de se+de te’ ambiguity, as shown in (2).

(1) McCain told Palin to vote for herself.

McCain to Palin: “You should vote for yourself.”
McCain to Palin: “You should vote for this woman.”

(2) McCain talked to Palin about advertising themselves.

McCain to Palin: “Let’s advertise ourselves, you and I.”
McCain to Palin: “Let’s advertise these guys.”

For the semantic approach to work here, we need to assume that the presuppositions of *tell* (singular *tell*) and *talk-to* (plural *tell*) are as follows (where ‘⊕’ is the operation that creates pluralities of individuals from singularities/pluralities of individuals). Below we give just the basic definitions of the verbs (their ‘de re’ counterparts work on a par with (23) in section 3).

(3) $[[\text{tell}]]^{\circ}(w)(y)(f)(x)$ is defined only if for all <w’, x’, y’> ∈ Dox$^{SG}$(w, x, y), f(w’)(x’)(y’) is defined (where f is of type <$s,<$e,<$e,t>>>).
(DoxSG(w, x, y) := {<w′, x′, y′>: x, who is addressing y in w, doesn’t rule out the possibility that he/she is living in w′, that he/she is x′ (i.e., x′ is ‘self’), and that he/she is addressing y′ (i.e., y is ‘addressee-self’) in w′}.)

(4) \([\text{[talk-to}^\text{pl}]^{σ}f^w(y)(f)(x)]\) is defined only if for all <w′, x′, x′⊕y′> ∈ Doxpl(w, x, y), f(w′)(x′)(x′⊕y′) is defined.

(Doxpl(w, x, y) := {<w′, x′, x′⊕y′>: x, who is addressing y in w, doesn’t rule out the possibility that he/she is living in w′, that he/she is x′, and the he/she is addressing y′ in w′}.)

Thus we arrive at the following verbal typology:

(5)  
   a. ‘Self’ quantifying verbs: e.g., Subject-Control promise and expect (as in Obama expects to outsmart himself); Object-Control convince (as in McCain convinced Palin to vote for herself); non-Control promise, convince and expect (as in McCain expects that he will outsmart himself).
   
b. ‘Addressee-self’ quantifying verbs: e.g., Object Control tell (as in McCain told Palin to vote for herself); non-Control tell (as in McCain told Palin that she voted for herself).
   
c. ‘Self+addressee-self’ quantifying verbs: e.g., Subject+Object Control talk-to PL (as in McCain talked to Palin about advertising themselves); non-Control talk-to PL (as in McCain talked to Palin about them advertising themselves).

We also need to assume that some verbs introduce two individual-binding indices (one corresponding to the ‘self’, and one corresponding to the ‘addressee-self’). And we need to make a change in the relevant part of the definition of Type-II covaluation in (24) (see Section 3).

(6) For any root node α and any attitude verb ATT such that α dominates ATT, let: (i) w(ATT) be the pronominal world-argument of ATT and Subj(ATT) be the external argument of ATT, (ii) S(ATT) be the first node that dominates ATT and all its syntactic arguments, (iii) R(ATT) be the highest node in the clausal complement of ATT that is c-commanded by all abstraction indices introduced by ATT, and (iv) k(ATT) be the world-binding abstraction
index introduced by ATT, and \( m(ATT) \) be the first individual abstraction index introduced by ATT. Then –

Any two NPs (of type e), \( \beta \) and \( \gamma \), that are dominated by root \( \alpha \) are Type-II covalued in \( \alpha \) if and only if there is at least one ordered pair \( \langle NP1, NP2 \rangle \in \{ \langle \beta, \gamma \rangle, \langle \gamma, \beta \rangle \} \) such that for any context \( c' \) such that \( g_c \) is appropriate for \( S(ATT) \) and \( [\[ S(ATT) \]]^{c'} \) is defined, and any \( g' \supseteq g_c \) such that \( \text{Dom}(g') = \text{Dom}(g_c) \cup \{ j: j \text{ is a pronominal index that is free in } NP1 \} \) and \( [\[ NP1 \]]^{c', g'} \) is defined:

(i) if ATT is ‘self’ quantifying, then for any \( <w, x> \in \text{Dox}(\[[w(ATT)]^{c', g'}, [\[ NP1 \]]^{c', g'}]) \):

\[
\{g'' \supseteq g': [\[ R(ATT) \]]^{c', g''} \text{ is defined} \} \cap \{g'' \supseteq g'[_{(\text{ATT})} \rightarrow w]; [\[ NP2 \]]^{c', g''} = x \} \neq \emptyset.
\]

(ii) if ATT is ‘addressee-self’ quantifying, then for any \( <w, x, y> \in \text{Dox}^{\text{SG}}(\[[w(ATT)]^{c', g'}, [\[ Subj(ATT) \]]^{c', g'}, [\[ NP1 \]]^{c', g'}]) \):

\[
\{g'' \supseteq g': [\[ R(ATT) \]]^{c', g''} \text{ is defined} \} \cap \{g'' \supseteq g'[_{(\text{ATT})} \rightarrow w, m(\text{ATT}) \rightarrow x]; [\[ NP2 \]]^{c', g''} = y \} \neq \emptyset.
\]

(iii) if ATT is ‘self+addressee-self’ quantifying, then for any \( <w, x, x \oplus y> \in \text{Dox}^{\text{PL}}(\[[w(ATT)]^{c', g'}, [\[ Subj(ATT) \]]^{c', g'}, [\[ NP1 \]]^{c', g'}]) \):

\[
\{g'' \supseteq g': [\[ R(ATT) \]]^{c', g''} \text{ is defined} \} \cap \{g'' \supseteq g'[_{(\text{ATT})} \rightarrow w, m(\text{ATT}) \rightarrow x]; [\[ NP2 \]]^{c', g''} = x \oplus y \} \neq \emptyset.
\]

\textit{McCain told Palin to vote for herself} is thus ruled in, because the clausal complement is always defined in \( \text{Dox}^{\text{SG}} \) of McCain (whenever the higher clause is defined), so \( \text{PRO} \) corresponds to whoever McCain thinks he is talking to (‘addressee-self’), and \textit{herself} denotes who he is actually talking to (in other words \( \text{PRO} \) and \textit{herself} are Type-II covalued by clause (ii) of (6)). But \textit{McCain told Palin to vote for himself} will be ruled out because the presuppositions of \textit{tell} only apply to McCain, not to Palin: there are bound to be contexts \( c' \) where the clausal complement is not defined in \( \text{Dox}^{\text{SG}} \) of Palin (in \( c' \)) – that is to say, there are bound to be contexts where Palin doesn’t entertain the presuppositions of the clausal complement. Therefore, in this case \( \text{PRO} \) and \textit{himself} are not Type-II covalued. \textit{McCain talked to Palin about advertising themselves} works in a similar fashion (except that this time, clause (iii) of (6) is the relevant one).

\textbf{Acknowledgments}
The idea for this paper came out of my discussions with Yosef Grodzinsky on acceptable “violations” of Condition B, and I thank him for his invaluable feedback. I am also indebted to Ilana Arbel, Gideon Avrahami, Gennaro Chierchia, Danny Fox, Jon Gajewski, Mira Goral, Irene Heim, Brian Leahy, Diane Lillo-Martin, Keir Moulton, Orin Percus, Alan Prince, Jim Pryor, Ken Safir, Philippe Schlenker, Roger Schwarzschild and William Snyder for very helpful comments, discussion and linguistic judgments. Most of all, I am grateful to Simon Charlow for sharing with me his work and thoughts on unexpected BT effects, to Jane Grimshaw for rescuing this project from a premature death, and to Moshe Sharvit, whose sound advice and common sense helped me make some non-trivial decisions. I alone am responsible for any errors in judgment – linguistics-related or other.

YAEL SHARVIT
Department of Linguistics
337 Mansfield Road
Unit 1145
University of Connecticut
Storrs, CT 06269-1145
e-mail: ysharvit@gmail.com

References


Heim, Irene. 1994a. Puzzling reflexive pronouns in de se reports. Unpublished handout (Presented at Bielefeld), MIT.


Heim, I. 2005. Features on bound pronouns. Ms., MIT.


Sharvit, Yael. 2009. Covaluation and Unexpected BT Effects. Ms, University of Connecticut, Storrs, CT.

