Direct Compositionality and Variable-Free Semantics: The Case of “Principle B” Effects*

PAULINE JACOBSON

6.1 Goals

This study starts from the premise that the hypothesis of direct compositionality (DC) is the simplest among current competing hypotheses as to the organization of the grammar, and so should be rejected only in the face of strong evidence to the contrary. The plot of this paper, therefore, is as follows. First I discuss the fact that under at least a reasonably strong version of direct compositionality, we would not expect to find constraints on non (strictly) local chunks of representations\(^1\) (especially constraints which ultimately affect the possible range of interpretations of some sentences). For under a strong version of direct compositionality, the grammar has no way to state such constraints. And yet, there is a large body of lore which takes it as axiomatic that the grammar does indeed contain such constraints and

\* The basic idea for Principle B which I pursue here came as the result of a conversation with Daniel Büring and Chris Barker on a lovely afternoon walk on a lovely beach in San Diego. When I got home and tried to reconstruct an idea Daniel had floated, it came out as roughly this proposal. But Daniel claims this isn’t quite what he had in mind, so I’ve been presenting this as my work. Nonetheless, Daniel deserves far more than the usual thanks, as also does Chris. In addition I’d like to thank audiences at the Direct Compositionality Workshop, at UConn, NYU, Harvard, and ZAS for helpful comments, especially Anna Szabolcsi. Finally, special thanks to Danny Fox whose discussant comments at the workshop raised a number of potential problems and areas that need further work—only some of which I have space to discuss here.

\(^1\) By “non strictly local” I mean a constraint whose domain is more than just a mother and its daughters. It should be noted that even a strictly local constraint on representations is not compatible with the direct compositional program (at least not with a strong version of it), for under this view one would not expect to find any constraints on representations at all. However, any strictly local constraint could be recast as a constraint on the syntactic combinatory rules which sanction the local bit of representation (e.g. phrase structure rules or their analog), and so it is only the non-local constraints that should worry a direct compositionalist.
this lore thus provides an obvious challenge to the (strong) direct compositional program. After all, we all learn in our introductory textbooks about "binding theory" principles such as A, B, C, and Weak Crossover; principles which are almost always stated in configurational terms and which make reference to the position of co-indexed material in global chunks of representation. There is also a second way in which "binding" phenomena provide a challenge to direct compositionality: we often find effects that look roughly like binding theory effects but where the appropriate representation on which to state the constraint is not found on the (audible) surface. Enter the notions of abstract levels of representation (e.g. LF) and/or silent or deleted material—which allow us to posit that the relevant representation is indeed found if one looks hard enough. But again these tools are generally incompatible with at least a reasonably strong version of direct compositionality, and so once again the apparent necessity for configurational constraints would seem to doom the direct compositional hypothesis. In view of all of this, it has become essentially a non-negotiable premise among a large number of linguists that DC is wrong (indeed, the way many textbook discussions proceed, the hypothesis of DC is not even on the radar scope).

But I propose to take a different strategy. Let us start with the observation that DC should be abandoned only in the face of strong evidence to the contrary. Is it really so obvious that the effects attributed to binding theory require statements about co-indexation which hold for non-strictly local syntactic chunks of representation? My claim is: no—quite the contrary. In fact, the usual strategy of casting these in this way is at best stipulative and unilluminating. At worst, the machinery that this strategy requires in order to really work is so baroque as to make this strategy completely implausible. This paper will be a case study intended to back up these claims: the case study centers on "Principle B" effects. After elaborating on these remarks in §§6.2–6.3 (and developing the tools of variable-free semantics which are crucial here to the direct compositional account that I will propose) the paper divides into two parts. Section 6.4 will document just how complicated it is to try to attribute these effects to a representational constraint on indices. Section 6.5 proposes an alternative account of these effects—one stated in the terms of a direct compositional (and variable-free) theory.²

² A different non-syntactic and non-representational account is explored in recent work by Schlenker (2005). Space precludes a comparison here with his account, although note that his analysis is not implemented in a direct compositional architecture.
6.2 The Hypothesis of Direct Compositionality and the Role of Structured Representations

In the broadest sense, the hypothesis of direct compositionality is compatible with a number of different theories of just how the syntax works and, in particular, a number of different possibilities as to how rich are the representations which are available to syntactic rules. (For some early discussion of this point showing that direct compositionality could be embedded within a full transformational framework, see Partee 1976; for more recent discussion of this general issue see Jacobson 2002.) That said, let me begin with a very strong version of direct compositionality—one which couples this hypothesis with a very constrained view of the syntax.

We can see any linguistic expression as a triple of (sound, syntactic category, meaning). Assume that a linguistic expression is nothing else: in particular it is not also associated with some kind of structured representation. Assume further that the grammar is simply a set of rules/principles (whichever label one prefers) which map one or more such triple(s) into another triple. (Note that the question of whether or not this is correct is quite independent of the question of whether or not the grammar contains just a few highly general statements.) Under this view, there is no room to state constraints on structured representations. For “structure” is not something that the grammar ever gets to see—indeed it is just a representation for the convenience of the linguist. A tree is merely a representation of the proof of the well-formedness of some string, and a very partial representation of how the semantics worked to put meanings together. And—a point well stressed within Generalized Phrase Structure Grammar (GPSG, see, e.g., Gazdar et al. 1985)—all constraints and phenomena must be encoded into the rules themselves (which again might be stated in highly schematic form) rather than being constraints on representations. An important point to keep in mind is that any theory needs combinatory rules (or “principles”) which “build” larger expressions from smaller ones—and so a theory which puts all of the work into these is adding no new machinery. The view that the grammar itself also keeps track of

---

3 Thus much of the work within GPSG and Categorial Grammar maintains a direct compositional architecture of the type laid out above, and at the same time uses only a few very general statements concerning the syntactic and the semantic combinatorics. Note, moreover, that both of these theories contain principles to predict the semantic combinatorics in terms of the syntax. I point this out here because the notion of “type-driven” interpretation originated within these theories (see especially Klein and Sag 1985) and yet has come more recently to be associated with the non-direct compositional architecture of a theory in which the syntax first produces representations (surface structures which are then mapped to LFs) which are subsequently interpreted by the semantics.
representations and uses these in the statements of other constraints requires extra machinery—and so the burden of proof should be on that position.

Of course it has been known since at least the mid 1980s—when there was vigorous discussion on the feasibility of the theory of GPSG—that natural languages require more than a context-free grammar (see, e.g. Culy 1985; Shieber 1985). I will therefore back off from the strongest version of direct compositionality given above, and will assume that grammars also contain Wrap operations (see, e.g. Bach 1979, 1980; Dowty 1982; Pollard 1984; Jacobson 1987 and others) which allow one string to be infixed within another. This means that it is not true that linguistic expressions have no associated structural information. For if there are operations which infix one expression into another, then there must at least be enough information to define the infixation point. Hence assume that the input to the rules are not pure phonological strings, but rather strings with a defined infixation point. (For some different ways to formalize Wrap operations, see Pollard 1984; Vijay-Shanker et al. 1986; Hoeksema and Janda 1988; Jacobson 1992). But we will tentatively adopt the view that this is the only amount of structure that the grammar gets to “see”: full-blown representations like trees are not part of grammar but again are just convenient representations for the linguist. Note, incidentally, that the class of languages which can be described by grammars with Wrap operations is—at least under some formalizations of Wrap—quite well understood (see, e.g. Vijay-Shanker et al. 1986). The known deficiencies with context-free grammars disappear, while there remain constraints on the class of possible languages allowed by this apparatus.

6.3 Relevance of “Binding” Phenomena

6.3.1 The Standard View

One of the most visible challenges to the DC view comes from so-called “binding theory” principles: that is, principles governing the distribution and interpretation of pronouns, reflexives, etc. In many textbooks and in many circles, there is an unquestioned assumption that in order to capture the distribution and interpretation of these items, the grammar must contain principles which constrain co-indexation and which are stated across non-local chunks of representation. Note that co-indexation is a purely syntactic notion and so such constraints in and of themselves would tell us nothing about the interpretation (which is ultimately what is at issue), but it is also assumed that these indices have an effect on the interpretation. Of course this program is also generally implemented within a theory with a level of LF
(which inputs the compositional, model-theoretic semantics), and so when I use the term "standard view" in this paper I generally refer to this conception of things.

Although this paper is primarily about Principle B, here I will illustrate the basic point with reference to another representationally stated constraint on "binding": the purported constraint that a "binder" must c-command a bindee. As point of departure for the discussion, then, consider (1) which is one possible way to state such a constraint:

(1) If a "binds" β (a notion to be defined below but which is assumed to hold at LF) and a is co-indexed with β, then a must c-command β.

Before continuing, let me clarify that the claim that a "binder" must c-command its bindee has been used in two different senses. In one sense, this is often claimed to be something which must hold at LF. But, as we will see below, this is a somewhat uninteresting principle, since once we give a sensible definition of "binding" this becomes true by definition (given a few completely standard assumptions). Thus (1) would amount simply to an observation (and a fairly trivial one) rather than a principle in the grammar. The second way in which this constraint is thought of is as a constraint on some non-LF level of syntactic representation—for the purposes of this discussion let us say surface structure. Under this view, the idea is that such a constraint is motivated by the existence of Weak Crossover effects, as in (2):

(2) (a) Every man; loves his; mother.
(b) *His; mother loves every man;.

(Here and throughout this paper I reserve subscripts like i and j as a way to notate the reading of a sentence without any commitment to these standing for indices used in the grammar. For indices themselves, I will use arbitrarily chosen integers or, when more convenient, I will use variables like m and n as variables over grammatical indices. I will of course ultimately be arguing that there are no such things as indices in the grammar.) I realize that there are any number of proposals as to how to account for WCO effects using all sorts of different theories and theoretical apparatus, so my discussion below of the defects in the formulation in (1) should not be taken in and of itself to constitute an argument against configurational constraints and/or against LF. Nonetheless, I think that exposing the problems in a statement like (1) is instructive of the types of problems which emerge with constraints of this nature.

Note that (1) involves three crucial ingredients: c-command, co-indexation, and, quite crucially, the notion that every man is in some special "binding"
relationship with *his*. What I would like to argue here is that there really is no simple notion of “binding” which takes care of the relevant cases. I will illustrate the point via a modified version of the account in Heim and Kratzer (1998). (The modifications here are almost all terminological; there is one content modification made for expository convenience (and which will be pointed out where relevant), but this actually has no effect on the central point here.) My reason for taking the HK account as a point of departure is simply because it is one of the few accounts which is fully explicit and which actually gives a definition of the semantic notion of “binding” and ties this definition into its use in the statement of “binding constraints”.

First, let us consider one way to view the LF for (2a), which is to assume that a surface sentence such as (2a) pairs with an LF of the form in (3).4

I am using the HK notation here (with my own addition of Λ as a node label), but a comment is in order about HK’s use of an index like 8 as a terminal symbol in the tree. The rationale for this is its role in the semantic composition. The system is such that the sister to this node will have as its meaning a function from assignment functions to propositions, and since it contains at least one “open variable *n*”, that means that the [[S]] in question will (in general) have a different value according to assignment functions which differ on the value that they assign to *n*. (More precisely: [[S]] is a function from assignment functions to propositions. The value of this function will (in general) be different for different assignment functions which disagree on the value assigned to *n*.) The role of an index sister to S in this system is to trigger “Λ-abstraction” and “close off” the variable *n*. Thus the interpretation of the expression I have labeled Λ is a function from assignment functions to properties, such that for any assignment function *g*, [[Λ]](g) is a function

---

4 When I discuss the HK system—and other systems within the basic framework making use of LF—I will use the term DP; when I switch to a discussion of this within CG I use the term “NP” for the same creature. It is hoped that this will cause no confusion; my rationale for this is to use the vocabulary which is standard in each theory.
which maps each individual $x$ into the value of $S$ on the assignment function just like $g$ except that $8$ is assigned to $x$.

A point which will become important below is that the use of the index as a terminal symbol in the tree is just one particular way to make sure that the meaning of $S$ (a non-constant function from assignment functions to propositions) "shifts" into a function from assignment functions to meanings of type $<e,t>$ by what we can think of as the semantics of $\lambda$-abstraction. One could have removed this label from the tree and posited a more complex combinatory rule for combining $S$ directly with the meaning of the DP. This is essentially the route taken in Montague (1973) (although his was framed in a direct compositional architecture where LF played no role in the grammar). Under this account a sentence roughly like that shown above combined directly in the syntax with the DP (which was substituted on to a pronoun in subject position), and the semantics involved two steps: shifting the meaning of $S$ by $\lambda$-abstracting over the relevant variable, and taking the resultant property as an argument of the DP. A third alternative would be to dispense with the "silent" operator $8$ here in the tree and leave the rest as shown there, with the node $\Lambda$ exhaustively dominating $S_8$ (hence note that the index is instead made a subscript on $S$); the rule interpreting $\Lambda$ on the basis of $[[S_n]]$ is the obvious one. I belabor these points here because my claim below is that once one considers the semantics, there really is no obvious sense in which every man "binds" his. I will develop this point using the HK details, but it is important to note that the point is independent of those particular details.

As noted above, this account of binding—and indeed any account which makes use of variables—posits that each constituent (at the level at which meaning is assigned) has as its semantic value a function from the set of assignment functions to something else. Assume further that each assignment function is a total function from the set of variables to model-theoretic objects (i.e. individuals in the cases relevant here). This is relatively standard but is actually different from the treatment in HK: they take assignment functions to be partial and I will comment on one reason for this below. However, since the exposition is eased by using total functions, we will do so here. Notice then, that if a constituent $C$ contains no unbound variables within it, it will denote a constant function from the set of assignment functions.

Let $G$ be the set of assignment functions, and let $G/n$ be any subset of $G$ such that each member of $G/n$ agrees on the value of all of the variables except that assigned to $n$. Then we can define the notion of a "free variable" as follows: $n$ is free within $C$ if there is a $G/n$ with two members $g_1$ and $g_2$ such that $[[C]](g_1) \neq [[C]](g_2)$. (A notational point: I use $[[C]](g)$ rather than the more familiar $[[C]]^g$ to highlight the fact that $[[C]]$ is actually a function
from the set of assignment functions and $g$ is an argument of that function.)
Informally, the idea is that a variable $n$ is free within some expression if $[[C]]$ gives
different values according to different ways to assign individuals to the
variable $n$. Otherwise, $n$ is *bound*. Once $n$ is "bound", then $[[C]]$ will assign the
same value to all assignment functions which differ only on the value that they
assign to $n$. A caveat here: strictly speaking, this is probably not the definition
of "free" that corresponds to our syntactic intuitions about what is free. The
reason is that there might be variables we would want to define as free within
$C$ but where $[[C]]$ is a constant function for irrelevant reasons, as for example
the case of tautologies like *8 walks or 8 doesn't walk*. This is avoided under the
partial function approach. Hence HK's definition of "free within $C" is actually
slightly different than that given here. But this modification will not affect the
main point below as the interested reader can verify.

So far, this gives a semantic basis for the notion of "free" vs. "bound", but
what does it mean to say that one thing "binds" another? HK's definition
proceeds in two steps, the first of which is to define "sem-binding":

**Definition of Sem-Binding:** $a$ sem-binds $X$ (for $X$ a pronoun or a trace) iff the sister to
$a$ is the smallest subtree in which $X$ is free.

Notice that this definition requires simultaneous reference to a *representation*
(since it uses the notion "sister", "subtree", etc.) and to the *interpretation* of that
representation. It is, then, a definition which can be checked by the grammar
only at the stage of a compositional interpretation of LF. If it is part of a
constraint (such as WCO) used in the grammar, the constraint itself must
simultaneously check the *level at which c-command is required to hold* (say,
surface structure), an *LF representation*, and the *interpretation of that LF.*

This already should make us suspicious that something is amiss: it requires
a very complex view of the architecture of the grammar. But there is another
problem: the above definition does not actually give us the notion of "binding"
that we need in order to account for WCO effects (or for other kinds of
effects which are often stated in terms of co-indexation and "binding"). For
according to this definition, the sem-binder for *his* in (3) is not *every man* but
rather is the node whose label is the index 8. Quite importantly, this is not just
some accidental property of this particular system or some accidental quirky
oversight on the part of HK. As long as *every man* is treated as a generalized
quantifier, then in all well-understood ways to do the compositional semantics
using variables there really is no privileged relationship between the semantic
contribution of the pronoun *hen* and *every man*. The only sensible way to
think about a *hen* as going from "free" to "bound" is to point to the step in
the semantic composition when the meaning changes from a non-constant
function on any G/n to a constant function on G/n. But this has nothing to do with the semantic contribution of every man.\footnote{In the HK system—in which assignment functions are partial functions—the "binding" of a variable is not a matter of going from non-constant to constant function, but is instead a change in the set of assignment functions for which the value of the expression is defined. But the remark above holds here too: this semantic shift again has nothing to do with the semantic contribution of the generalized quantifier.} It can be done by a silent operator given as an integer (as in HK), by one of the steps in Montague's "Quantifying In" treatment, or by a type-shift rule. In first order logic (or in other systems where quantification is over assignment functions)\footnote{As, for example, is argued for in Heim (1997); there it does make some sense to talk about the quantifier as "binding" the pronoun.} it makes some sense to think of the contribution of the quantifier as "binding" the pronoun, but not in any system where every man is of type \(\langle\langle e, t\rangle, t\rangle\).

Therefore, in order to establish any relationship between the DP and the pronoun, HK define a second notion of binding based on the first. They call this a "derivative notion of semantically bound"; I will refer to it as "LF binding":

**Definition of LF binding:** \(a\) LF binds \(\beta\) iff \(a\) and the trace of \(a\) are sem-bound by the same thing. (Note: HK refer to this as a "derivative notion of semantically bound")

I can now cash in a promissory note given earlier: the claim that a "binder" must c-command its "bindee" at LF now follows by definition (as long as we also assume that a DP must c-command its trace at LF). The interested reader can verify for themselves that this follows from the definitions here; it will follow equally well in most other systems making use of this type of apparatus.

Of course in standard accounts this definition of "binding" is supposed to interact with constraints on indices in the syntax (this is the whole point, after all—this is used to get typical "binding theory" constraints). Hence in order to tie this definition into something of use to the syntax, HK first give a further definition:

**Definition of syn(tactic) binding:** \(a\) syn-binds \(\beta\) iff
\[
\begin{align*}
& a \text{ is co-indexed with } \beta \\
& a \text{ c-commands } \beta \text{ (at some level—assume here surface structure)} \\
& \ldots \text{ (other conditions here would be to capture other aspects of "binding")}
\end{align*}
\]

Finally, WCO—and other bits of the "binding theory"—will now follow from the constraint in (4) (notice that this replaces (1)):

\[
\text{(4) If } a \text{ LF binds } \beta \text{ then } a \text{ must syn-bind } \beta \text{ and vice versa.}
\]

It seems to me that this account leaves plenty of room for suspicion. As pointed out above, the relationship between co-indexation and true semantic...
"binding" is both complex and indirect, and it took considerable work to come up with a definition of "binding" according to which every man "binds" the pronoun. And, as already discussed above, a constraint such as (4) requires reference to multiple levels simultaneously: surface structure (or whatever is the level at which "syn-binds" is intended to hold), LF representation, and the interpretation of LF. Of course this is only one version of WCO—a defender of this basic program might hope that the fault lies not with the general strategy but with this kind of account of WCO. I suspect, though, that many different accounts—both of WCO and of other kinds of constraints on "binding"—will suffer from similar problems. Perhaps we are on the wrong track trying to reduce the empirical effects to constraints which make use of configurational properties. The effects are, after all, semantic effects—can they be built directly into the semantic (and possibly also syntactic) combinatorics as opposed to building them into representations? I will argue below that they can.

But before going there, let me point to another reason to be suspicious of configurational accounts. The fact is that we very often find cases which show "binding" types of effects but where the relevant configuration is not found in the audible (or visible) surface structure. Typical cases which have been discussed over and over in the literature involve a host of "connectivity" effects in, for example, copular sentences. Of relevance to the point here would be "bound variable connectivity" as in (5):

(5) The woman who every Englishman loves (the most) is his mother.

Since it is so often assumed that a "binder" must c-command a pronoun that it "binds", it is assumed that the post-copular constituent here must (at some level) be a full sentence every Englishman loves his mother. Of course, unless one posits silent material surrounding his mother at surface structure, the level at which the c-command condition is met is not surface structure, so one would need to define the WCO requirement as holding for some other level. Incidentally, some of the literature which invokes a "c-command" constraint on binding as evidence for additional material surrounding his mother seems to be concerned not with the fact that there would otherwise be a WCO violation, but rather seems to assume that a "binder" must c-command a "bindee" at LF (hence, this literature posits that at least at LF we have material surrounding his mother). But as we have seen above, there is no empirical content to the claim that a "binder" must c-command its "bindee" at LF—it is true by definition. Thus I assume that what people really mean when they say that there is a c-command constraint at LF and that hence every Englishman must c-command his at LF is that there is no other way to
give a semantics for this sentence. But first, simply positing a representation with c-command is not the same as actually showing that this does provide a semantics for this sentence.7 Second, as shown in Jacobson (1994) (see also Sharvit 1999), the claim that there is no way to give a semantics for (5) without having his be c-commanded by a “binder” is incorrect: the correct meaning can easily be compositionally composed (without c-command) by a perfectly natural extension of the analysis of functional questions (Groenendijk and Stokhof 1983; Engdahl 1986). Jacobson (1994) further shows that this comes for free under the variable-free program, and nothing extra is needed to get a meaning for this sentence without positing silent, deleted, and/or reconstructed material in the post-copular position. (We return to this briefly in §6.5.3.) The moral: the belief that there are configurational constraints on indices (which play an indirect role in getting the semantics right) forces the positing of extra material in (5) and hence extra machinery to make sure that it is silent. All of this is simply to satisfy configurational constraints which are quite complicated to state in the first place. Thus I hope to have convinced the reader that perhaps it is best to move on to an entirely different strategy.

6.3.2 The Variable-Free Alternative
To elucidate a non-configurational alternative, I will first briefly review the variable-free approach to pronoun “binding” discussed in detail in Jacobson (1999, 2000). By way of comparison, consider the following properties of the standard view. Take sentences such as:

(6) (a) Every mani believes that heii lost.
(b) Every mani loves hisn mother.

Under the standard view, any expression which is or contains within it an “unbound” pronoun is (by definition of “unbound” as discussed above) a non-constant function from assignment functions to propositions. Hence [[heii lost]] is a (non-constant) function from assignment functions to propositions, [[hisn mother]] is a (non-constant) function from assignment functions to individuals, and [[heii]] is a (non-constant) function from assignment functions to individuals, where [[heii]](g) = g(n). (Note again that I use i in

7 To be fair, there is one proposal which actually does give a semantics for the associated representation: this is the “question/answer” proposal in Ross (1972), recently revived in den Dikken et al. (2000) and Schlenker (2003) (see also Romero (this volume)). A discussion of the feasibility of this proposal is obviously beyond the scope of this paper; for some relevant discussion see Romero (this volume) and Caponigro and Heller (this volume).
the above sentences to simply illustrate the relevant meaning, whereas \( n \) is a variable over an actual index.)

Under the variable-free view put forth in Jacobson (1999, 2000), an expression with a pronoun which we intuitively would want to call “unbound” within that expression denotes a function from individuals to something else. Hence *he lost* denotes a function from individuals to propositions; *his mother* is a function from individuals to individuals, and so is *he* (where the particular function denoted by *he* is the identity function—modulo gender information which I will ignore here). There are no indices here, and so we see right away one interesting difference between this and the standard view. Under the standard view there are actually an infinite number of (homophonous) pronouns, each with a different index: under the variable-free view there is but one (ignoring case, number, and gender).

Thus *he lost* is the function which maps each individual \( x \) into the proposition that \( x \) lost. I will continue to ignore the contribution of gender, and with this simplification, *he lost* thus has the same semantic value as *lost*. How do we get this result? Phrasing this more generally: pronouns occur wherever NPs can occur. If an NP is of type \( e \), the function which takes it as argument has as its domain things of type \( e \). But then how can a pronoun—or material which contains a pronoun unbound within it—occur in the same places, since these have meanings of type \(<e,e>\)?

The answer is via a unary rule; that is, a rule taking a single expression (a triple) as input and returning a single expression (a triple) as output. “Type-shift” rules are a subspecies of these rules but are not the only logically possible ones. In fact, most of the work on “type-shift” operations is framed in a non Categorial Grammar framework and is often seen as a rule which shifts only the meaning (and the semantic type) and not also the syntactic category. But since I am embedding the general program within a Categorial Grammar syntax, any unary rule which shifts the semantic type of something will also have to shift its syntactic category. I assume basic familiarity with the Categorial Grammar notation; one addition to the usual CG apparatus is the existence of categories of the form \( A^B \) (for any two categories \( A \) and \( B \)). Any expression whose category is \( A^B \) denotes a function from \( B \) type meanings to \( A \) type meanings. Semantically, then, these would be the same as a category of the form \( A/B \) but the syntax is different: an \( A^B \) is something that generally distributes like an \( A \) but contains within it (or is) an (unbound) proform of category \( B \). A pronoun is thus listed in the lexicon as having category \( NPNP \) and its meaning is the identity function on individuals. We can now formulate a rule which allows an ordinary expression such as *lost*—which wants an NP in subject position and is a function of type \( <e,t> \)—to map into a new
homophonic expression wanting an NPNP in subject position and wanting an argument of type <e,e>. I call this the “Geach” rule (Geach 1972) and notate this as g:

(7) Semantics of the \( g \) rule: let \( f \) be a function of type \(<a,b>\), then \( g(f) \) is a function of type \(<<c,a>,<c,b>>\), where \( g(f) = \lambda v_{\text{of type } c} [\lambda x_{\text{of type } c} [f(v(c))] \]

(Note that this is a unary version of the function composition operator: it takes a function \( f \) and maps it into a new function which takes as argument a function \( h \), such that \( g(f)(h) = f \circ h \). The full \( g \) rule, then, is formulated as in (8); I will hereafter use the prime-notation \( (a') \) interchangeably with the double bracket notation \( [[[a]]] \) to notate the meaning of \( a \):

(8) Let \( a \) be an expression of the form \(<[a]; A/B; a']\). Then there is an expression \( \beta \) of the form: \(<[a]; A/C/B^C; g(a')]\>

Example (9) illustrates the syntactic and semantic composition of \( \text{he lost} \):

(9) \text{lost; S}_{NP}/NP; \text{lost; } \rightarrow g\text{lost; } S^{NP}_{NP}/NP; \lambda f_{<e,e>}[\lambda x[\text{lost'}(f(x))]]
\text{he; NP}^{NP}; \lambda y[y]
\text{he lost; } S^{NP}; \lambda f_{<e,e>}[\lambda x[\text{lost'}(f(x))]](\lambda y[y]) = \text{lost'}

To complete the analysis of (6), we need one further unary rule which accomplishes the effect of “binding”. Of course I should point out that there is no real notion of “binding” in the sense of some notion which plays a role in the grammar, and so I use this term informally to mean the operation which is necessary to give a particular reading. Thus “binding” is accomplished by what I call the \( z \) rule. I first define its semantics in (10), and the unary rule is given in (11):

(10) Let \( h \) be a function of type \(<a,<e,b>>\). Then \( z(h) \) is a function of type \(<<e,a>,<e,b>>\) where \( z(h) = \lambda f_{\text{of type } <e,a>}[\lambda x[h(f(x))(x)] \]

(11) Let \( a \) be an expression of the form \(<[a]; (A/NP)/B; a']\). Then there is an expression \( \beta \) of the form \(<[a]; (A/NP)/B^{NP}; z(a')]\>

For a generalization of this to the case of three-place verbs see Jacobson (1999). (The fact that (11) is formulated to apply only to items of category \((A/NP)/B\) (rather than more generally to \((A/C)/B\) and the corresponding fact that \( z \) is defined only for functions which contain an e-argument slot is purely for exposition; there is as far as I know no reason not to give (11) in fully general form.) Hence the semantic composition of Every man \( i \) believes that he \( i \) lost involves \( z \) on believes as follows:
A case like every man loves his mother is similar. his mother denotes the function mapping each individual into that person’s mother (again, ignoring gender information); call this the-mother-of function. loves undergoes z to expect as object an expression containing a pronoun, and its new meaning is \( \lambda \text{x}[\text{loves}'(\text{f}(\text{x}))(\text{x})] \). When this combines with the object we get the syntactic expression loves his mother whose meaning is \( z(\text{love}') \) applied to the-mother-of function. That boils down to the set of self-mother’s-lovers, and that occurs as argument of the subject. For full details, see Jacobson (1999) in which it is shown that the system (with modest generalizations on the rules above) can handle cases with any number of pronouns, any number of binders, and bindings in any order.

Readers unfamiliar with direct compositionality and with Categorial Grammar style analyses are often uncomfortable with the fact that this approach to binding makes use of two unary rules. There are some observations which should set such readers at ease. First, one can always trade in a unary rule for a silent operator, if the latter gives greater comfort. So instead of having a rule shifting the meaning (and syntactic category), one could recast \( z \) and \( g \) as silent little lexical items which combine in the syntax with believe and love. (Any unary rule can be repackaged as an “empty operator”, and vice versa.) Second, note that every analysis (that I know of) involves a unary (i.e. type-shift) rule and/or an empty operator to perform binding. In the standard view it is the rule (or operator) which accomplishes \( \lambda \)-abstraction. The difference, then, is not between the existence of a type-shift rule (or operator) to do binding, but simply the domain to which this rule applies. In the standard view, the “binding” of the pronoun involves a shift which takes place on a big domain: the meaning of a sentential expression like \( t_8 \) loves his mother shifts (from open proposition to closed property). Here the shift is a very local one, and applies instead to loves in (6b) (and believe in (6a)). But both systems make use of a unary (type-shift) rule (or empty operator). In fact, then, there is just one additional unary rule here: the g rule. But this is a rather simple and natural rule; it is just a unary (Curry’ed) version of function composition.

Under the “Derived VP” approach to binding (Partee and Bach 1981), the domain is somewhat smaller; the meaning of the VP shifts.
Let me forestall one potential misunderstanding. I am certainly not claiming that there is no syntactic side to "binding" and to the distribution of pronouns, and I am not claiming that everything is "done in the semantics". In fact, such a claim would be somewhat at odds with the general Categorial Grammar program, in which the two systems are tightly linked. Indeed it is quite crucial that the system does make use of syntactic information: the syntactic categories play a role in the statement of the rules above and so the grammar most certainly does get to "see" syntactic category. (As would be the case in any system.) What it does not get to see, though, is configurational information.

Can we thus get the effects of supposed configurational constraints such as, for example, WCO? Indeed we can—and in fact it is built into the system above. "Binding" is the result of the z rule: this rule maps a function of type <a,<e,b>> into one which wants an <e,a> as its first argument, and it "merges" the newly created e-argument slot with one of its later argument slots. Syntactically, it allows for its first argument to contain a pronoun. Thus the open argument slot which is ultimately contributed by the meaning of the pronoun is "merged" with a higher argument slot, and this is what gives the right semantics. If we assume that z is the only rule that accomplishes "binding", then the WCO effects automatically follow. In order to get WCO violations such as that shown in (2b), we would need a backwards version of this rule (call it s) which mapped a function of type <e,<a,b>> into one of type <e,<<e,a>,b>> and "merged" the newly created e-argument slot (which will ultimately be contributed a pronoun) to the earlier (or, lower) argument position. (For details of how this could indeed give a WCO violation, see Jacobson 1999.) The bottom line, then, is that the effect is built into the combinatory system rather than being stated on a configuration which is the result of the combinatory system. Since any theory does need the combinatory rules (including unary rules) the hope is that all such effects can follow from the rules and no use of representations is needed. This is what DC would lead us to expect.

Two final points before concluding this section; one concerns the analysis of free pronouns, as in:

(13) He lost.

In the account here, this actually does not denote a proposition, but rather a function from individuals to propositions. I assume that in order to extract propositional information, a listener supplies this to some contextually salient individual. Note that the standard account making use of variables has no real advantage here. For here too (13) is not a proposition—but a function from
assignment functions to propositions. In fact, all sentences are functions from assignment functions to propositions. Yet listeners do compute propositional information. In the case of a closed sentence (with no unbound pronouns) there is no mystery as to how this is done: since these denote constant functions from assignment functions, it does not matter which assignment function is picked. But in (13), it does. Presumably, then, a listener computes propositional information by applying this to some contextually salient assignment function (as opposed to the tack taken here, where the function is applied to a contextually salient individual).

A second point of interest concerns the analysis of (5), which shows "bound variable connectivity" without apparent c-command. But this is quite unproblematic in this account—the notion that a "binder" c-commands a "bindee" is all an illusion. (In fact, there is no real notion of binding, binders, etc.—we merely have the semantic composition doing its job, and the z rule "merges" an argument slot contributed by a pronoun to a higher argument slot.) More concretely, the semantic composition of (5) can be shown informally as follows (see Jacobson 1994 for more thorough discussion):

(14) the unique function f with range woman such that every Englishman z(loves) f is the-mother-of function.

As discussed also in von Stechow (1990) and Sharvit (1999), this involves just a generalization of the analysis of functional questions given in Groenendijk and Stokhof (1983) and Engdahl (1986). But what is of interest here is that functional questions in general and extensions to these cases follow immediately from the mechanisms discussed above. Since loves can undergo z, it follows that there can be a "functional" gap here. These phenomena are not some special creatures requiring new sorts of traces or other devices—they are just part and parcel of the machinery needed for "binding" in general. Moreover, the fact that the post-copular constituent can denote the-mother-of function is an automatic consequence of the system here too (since it contains a pronoun, it is a function from individuals to individuals). I should note that this analysis of the compositional semantics of (5) is sometimes reported in the following terms: the functional analysis of (5) allows one to construct a semantics in which his is not c-commanded by every Englishman and is therefore not actually "bound" by every Englishman. This kind of phrasing misses the point. Under the variable-free view, "binding" (in the sense of some privileged semantic connection between every man and the pronoun) and "c-command" are always an illusion—and no new analysis of "binding" is needed for the case here.

The remarks above illustrated the direct compositional and variable-free hypotheses. But the ultimate success of these two related hypotheses depends
on their ability to meet three challenges: (i) *The challenge for direct compositionality:* can all phenomena which have appeared to necessitate non-local constraints on representation be handled (in a simple and natural way) without such constraints? (ii) *The challenge for variable-free semantics:* can all phenomena which have appeared to necessitate constraints on co-indexation (and hence indices) be accounted for in some other way (and in a simple and natural way)? (iii) *The challenge for the purely semantic view of connectivity:* can all connectivity effects be accounted for without abstract levels and/or silent material? Obviously I am not about to try to answer all of these questions here, but my strategy will be to investigate one persistent "thorn" for direct compositionality and variable-free semantics: Principle B effects. The remainder of this paper attempts to demonstrate two things. First, the account of Principle B effects using representationally stated constraints on co-indexing requires so much unmotivated complexity as to be extremely dubious. Second, these effects can indeed be handled (much more simply) under direct compositionality and variable-free semantics.

6.4 Principle B Effects: The "Standard" Account

Principle B is supposed for the contrasts shown below. I am surrounding these with some discourse context, since judgments about coreference possibilities are often highly context-dependent and some of the claims in the literature about the lack of possible coreference might well hold only because it is difficult to imagine appropriate contexts. I have therefore tried to give the (b) cases every possible chance to be good, and the contrast between (a) and (b) seems to show that the Principle B effect is quite real (note that all of the (b) cases become perfect when a reflexive replaces the pronoun):

(15) What happened at the press conference yesterday?
   (a) Bushi defended hisi decision to go to war.
   (b) *Bushi defended himi (in the face of severe criticism).

(16) How does Bush manage to fool so many people?
   (a) Well, hei praises hisi policies whenever he gets a chance.
   (b) *Well, hei praises himi whenever he gets a chance.

(17) What happens on campaign trails?
   (a) Oh, every candidatei goes around praising hisi mother (for raising him so right).
   (b) *Oh, every candidatei goes around praising himi.
And so, conventional wisdom tells us that there is a constraint such as the following:

(18) A pronoun cannot be co-indexed with a c-commanding NP within the same local domain.

The definition of "local domain" varies from account to account—for our purposes let us take the local domain of some node to be the lowest S or NP node dominating it—but the particular definition will have no effect on the points in this section.

Before continuing, let me note that there are a number of well-known complexities as to just where we find these effects. They are, for example, mitigated with focus on either the subject or on the object pronoun. Moreover, the strength of the Principle B effect varies with the verb in question. praise—which I use throughout—is perhaps not the strongest verb to illustrate the point (introduce displays extremely strong Principle B effects, as will be discussed later). I nonetheless continue to use praise because the effect here seems strong enough to demonstrate the point (introduce would involve the additional irrelevant complication of dealing with a three-place verb). I will return to some of the issues in the final section (see especially §6.6.2); for now I will flatten the domain in a way which, hopefully, does no significant damage to the theoretical points.

Thus the existence of a principle such as (18) creates obvious challenges to the claims here. First, of course, it is not stated as a strictly local constraint (on sisters and/or mothers) and so could not be directly recast as a constraint on the combinatorics. Second, its statement requires the use of indices. And third, Principle B displays the typical connectivity effects, as shown in (19):

(19) (a) *What Bush always does (on the campaign trail) is praise him.

(b) *What every candidate does (on the campaign trail) is praise him.

If one were to take the position that the post-copular constituent contains no deleted, reconstructed, or silent material, then (18) cannot account for the deviance here, for him is not co-indexed with a c-commanding NP in the relevant local domain.

6.4.1 The Inadequacies of a Constraint on Co-Indexation

6.4.1.1 Problem 1: Coreference with Free Pronouns Yet there are at least two well-known problems with the co-indexation account. First, it has been known since at least as early as Reinhart (1983) that a constraint on co-indexation is not good enough for cases such as (15b) and (16b). Given the usual assumptions about how "binding" works, it will block the relevant
reading of (17b)—where we have a bound pronoun—but it will not suffice to block the readings shown in (15b) and (16b). This is because in (15b), nothing would stop the pronoun from being a free pronoun, which happens to pick up the individual Bush who is, after all, quite contextually salient. And yet this reading is still impossible. Similarly for (16b): why can both pronouns not be free, not co-indexed, and yet both happen to pick up the same individual?

Interlude: Are we sure that this is a problem? Obviously my answer to the question raised above in the subsection title is going to be “yes”—but I think it is worth pausing to consider a possible “way out” for the Principle B account. Could we posit a constraint on assignment functions which would rule out the possibility of differently indexed pronouns (or an NP and a pronoun with different indices) picking out the same thing? The idea is that if the pronoun were free, the only way it could pick out the same individual as the subject is to have a different index on it. Hence a constraint on co-indexation (classic Principle B, as stated in (18)) will suffice: (18) rules out the co-indexed representation, and the constraint on assignment functions rules out the other possible way to get the “wrong” interpretation of this sentence. In fact I suspect that this is essentially what was assumed in the early literature on indices: the assumption there seems to have been that there is a one-to-one correspondence between individuals and indices, which made the idea of a constraint against co-indexing look like a reasonable way to account for constraints on coreferential interpretations. While standard views of assignment functions do not contain such a constraint, it is worth revisiting the issue to see if this can work.

First, however, we should be a bit more explicit about the LF representations for, for example, (15b); this will also come in useful later. We will assume that Bush can be in the raised position at LF and can bind a trace in subject position (this assumption is fairly standard in order to allow for the possibility of sloppy readings with VP Ellipsis; I come back to this below). Take any “QRed” DP. It will be the sister to some index n and there will be a tn in the position from which it raised. We can ask about the index on the DP itself in LF. Must it have the same index? Can it have a different index? Or will it have no index? (See, e.g., Büring 2005 for extensive discussion.) This point will become important in the section “Interlude again” below; for now we will simply not notate any index on it (as already shown in §6.3.1, the index does no work in any case). Thus (15b) can have two different LFs:

\[(20) \quad (a) \quad \text{Bush} [_{\Lambda}8 \ [t_8 \ \text{defended him}_8]] \]
\[(b) \quad \text{Bush} [_{\Lambda}8 \ [t_8 \ \text{defended him}_7]] \]
So the question at issue here is the following: can the representation in (20b) give rise to the “reading” (or “understanding”) shown in (15b)? The proposal under consideration here is that it cannot, if we constrain assignment functions as follows:

(21) For any assignment function g, if m ≠ n, then g(m) ≠ g(n)

The interpretation of free pronouns presumably proceeds as follows: a sentence with a “free” pronoun within it has as its value a non-constant function from assignment functions to propositions. Its interpretation involves the listener applying this to some contextually salient assignment function. But on any assignment function g, [[Λ]](g) is the function that maps an individual x to the proposition that x defended g(7), where g(7) cannot be x. So pick any g such that [[Λ]](g) is a function which is defined for Bush. It follows that under that assignment function g(7) cannot be Bush, and so there is no way to get the bad reading indicated in (15b). (Incidentally, my own account will bear some resemblance to this, although this will not be achieved through the use of assignment functions.)

One might be tempted to immediately dismiss the constraint in (21) on the grounds that it would not allow for the strict reading of (22):

(22) Georgei loves hisi mother, and Dick does too.

The usual way to think about this is that the sloppy reading comes about by having the first clause have an LF representation analogous to (20a) (where the pronoun and the subject trace are co-indexed), while the strict reading is the result of the representation analogous to (20b) (the pronoun is free and happens to pick up George). And so the constraint in (21) would seem to incorrectly rule out the strict reading. But, in fact, to dismiss (21) on these grounds would be premature, for there is another story that we could tell about the strict reading. The details depend on just how to treat VP ellipsis: to pick one approach, assume that VP ellipsis involves “supplying” a meaning of type <e,t> at the position of the ellipsis site, where that meaning must be the meaning of some other expression at LF. Then the sloppy reading can come about in just the way noted above, but for the strict reading we can simply be picking up the meaning of an LF VP—whose meaning is just the value of the expression loves B's mother. (Thus for the sloppy reading the missing meaning—on any assignment function g—is the set of self-mother lovers—which is the meaning for the entire constituent labeled Λ—and for the strict reading the meaning—on any assignment function g—is the set of individuals who love g(8)'s mother. In both cases consider only assignment functions in which g(8) = George, and we get both readings.)
Nonetheless, the constraint on assignment functions proposed in (21) is untenable: it gives rise to a class of problems which can be exemplified by the following:

(23) Bill thinks that no one at the moment is living in his house.

If we were to adopt (21), then we incorrectly predict that (23) means that Bill thinks that no one other than himself is living in his house. To demonstrate: assume for the moment that no one has to undergo QR here. Suppose that Bill (or, the trace of Bill) and no one have the same index. Then in that case we do not get the reading in (23) because no one will “bind” the pronoun. The LF will be:

(24) Bill [t8 thinks that no one [t8 is living in his house]]

So suppose instead that Bill (or its trace) and no one (or its trace) have different indices; for example let no one (or its trace) have the index 6. The relevant LF is then:

(25) Bill [t8 thinks that no one [6 [4; is living in his house]]]

This means that on an assignment function g, the argument of [[no one]](g) is the function which maps each individual x into the proposition g'(6) lives in Bill’s house where g' is the assignment function just like g except that x is assigned to 6. However, there is no g' just like g except where Bill is assigned to 6 (because Bill has already been assigned to 8). Hence the argument of no one has to be a partial function, undefined for Bill. Thus Bill must believe that there is no one in the set characterized by the (partial) function which maps an individual to true in case they live in Bill’s house and are not Bill—and so we get the reading where Bill believes that no one but himself is living in his house. It is quite clear that this is not the right reading for this sentence.

There is one last trick we might try to save this constraint on assignment functions. We began the discussion of (23) by assuming that no one had to undergo QR. But suppose that that is wrong: let it stay in the subject position where the meaning of the VP directly is taken as its argument in the semantic composition. The point of this “trick” is that in this case its index will not matter for anything (there is never a stage in the semantic composition at which we λ-abstract over the index of a variable in subject position). And so no one could be co-indexed with Bill but without the bad consequence seen earlier: it will not mistakenly “bind” the pronoun his. This final attempt is easy to sabotage, simply add in another pronoun which must be “bound” by no one, as in:

(26) Bill, thinks that no man could possibly live in his house unless he can stand cockroaches.
The interested reader can verify that—given the general way of doing the semantic composition under discussion here—*no man* would have to undergo QR and hence the bottom line will be that this should have an “except Bill” reading, which it does not.

Interestingly, my own proposed solution also results in a somewhat specialized version of the problem discussed here. But in the case of my proposal the “problem” (if it is one) appears only in a special set of cases (cases which exemplify the Principle B environment). As it turns out, in just those cases, in fact, it is not clear that the prediction that there is a “no one but self” reading is incorrect. The facts are a bit complex and so I return to this in §6.6.

6.4.1.2 Problem 2: Co-Bound Material A related problem with Principle B has been discussed (e.g. Heim 1993; Fox 2000; Büring 2005): the co-indexing constraint given in (18) does not rule out the indicated reading for a case such as (27). The reason is that other facts (to be discussed momentarily) force us to allow for an LF such as (28), where this would correspond to a surface structure in which the subject and object of *praise* are not co-indexed.

(27) *Every candidate*$_i$ *thinks that he$_l$ should say that he$_i$ will praise him$_i$.

(28) ![Diagram of syntactic structure](image)

*Interlude Again: Are we sure that LFs such as (28) are possible?* Once again, I am going to answer the question posed immediately above in the affirmative (of course this answer is relative to the hopefully counterfactual world in which there are indices and LFs). This question is actually addressed in past literature (see especially Fox 2000); I will take Fox’s discussion as a point
of departure but will fine-tune his argument slightly. The question at issue, then, is whether the index on the binder must always match the index which performs the semantics of λ-abstraction. In other words, is the following tree bit well-formed?9

(29)  he₈[₇  S]

If this were ill-formed, then Principle B (as stated in (18)) would be enough to block the indicated reading in (27); Principle B blocks the representation in which the subject (or subject trace) is co-indexed with the object pronoun, and (28)—which seems to be another way to let in the bad reading—would be ill-formed.

Fox (2000) claims that the well-formedness of (28) can be demonstrated by considering the range of possible interpretations for (30):

(30)  Every manᵢ only thinks that HEᵢ should give a present to hisᵢ mother.

The reading of interest here is:

(31)  every man only thinks that HE (bound) should give a present to his (bound, but non-sloppy) mother

(That is, each man wants to be the sole present-giver to his own mother.) Hence this reading can be associated with the LF in (32) which, note, allows for a case where the “LF binder” (by the HK definition earlier) of a trace has a different index than the trace itself:

(32)  every man₈ [₈ [₇ thinks HE₈ [₇ [₇ should give a present to his₈ mother]]]]

This makes the alternatives within the focused domain be things like:

(33)  he₆ [₇ [₇ should give a present to his₈ mother]]

₇ is bound all the way through. Hence the alternatives within the thoughts of each man are alternative present-givers (as we want), but the mother stays constant (as we want).

But we are not quite home free if we wish to show that (32) (and hence more generally the configuration in (29)) should be allowed. For we do not really need to resort to the representation in (32) to get the relevant meaning. Suppose we simply do not apply QR to HE, so that the relevant LF is:

9 One answer to this put forth in Fox (2000) is that such representations come “for free” and thus something extra would be needed to block them. But this is not really true. Suppose that we assume that all DPs (proper names, quantified DPs, etc.) have an index and that this index is preserved under QR. Then representations such as he₈[₇ t₇ say he₈ praise him₇] would simply never come about (I’ve underlined the offending portion). It is true that the indices on full NPs and quantified NPs do no work, but it’s easy enough to ensure that they match the index on the “binder”. (For relevant discussion, see Büring 2005)
(34) every man$_8$ [8 [t$_8$ thinks HE$_8$ should give a present to his$_8$ mother]]

since HE$_8$ is what is focused, the relevant alternatives are he$_7$, he$_9$, etc. and so we get alternatives such as:

(35) he$_9$ should give a present to his$_9$ mother

which is what we want. (The alternatives within each man’s thoughts are exactly the same in (32) as in (34) since the two are semantically equivalent.)

Nonetheless, the basic argument is correct and just needs a bit more work to get it to go through. We can construct more involved versions of this case where the way out in (34) is unavailable. The strategy is to make sure that HE must be QR’ed by having it bind a different pronoun, one which is interpreted sloppily. Consider, then, the scenario in (36), followed by the sentence in (37):

(36) Scenario: There was a bad accident involving a bus full of elderly women from a nursing home. Several were injured, and brought to the hospital. Of those, many had sons who came to the hospital to see them—and each woman had several other relatives who also came. As is typical in hospitals, all the sons and the relatives were waiting around for hours—not being able to see their mother/relative and not even being told when they would be able to see them. This of course was very frustrating.

(37) Each man$_i$ hoped that at least HE$_i$ would soon be told when he’d$_i$ be allowed to see his$_i$ mother.

The relevant (and very salient) reading is notated in (38):

(38) Each man hoped that at least HE (bound) would soon be told when he’d (bound, sloppy) be allowed to see his (bound, strict) mother

In other words, the alternatives within each man’s hopes are about alternative relevant people (hence, the other relatives of that man’s mother) who might be told something about when each of these alternative people might get to see someone, but the mother remains constant in each person’s thoughts. If HE were not raised here then the representation would be as in (39), and the alternatives (within each man’s hopes) would be about other people being told when he (the man himself) could see that man’s mother:

(39) every man$_8$ [8 [t$_8$ hoped HE$_8$ be told when he$_8$ could see his$_8$ mother]]

To get the relevant reading we need the representation in (40):

(40) every man$_8$ [8 [t$_8$ hoped HE$_8$[7 [t$_7$ told when he$_7$ could see his$_8$ mother]]]]
But in that case, we must allow the configuration in (29) which in turn means that Principle B (as stated in (18)) cannot block the “co-bound” reading for (27), since the LF in (28) cannot be ruled out.

There is still one further tack that one might take to allow LFs like (40) while still blocking the LF in (28). Note that HE_8 is focused. Suppose we assume that only focused pronouns can raise and thus have a different index from the index on the trace that they bind. This might be a tenable position—but we can show that one can construct cases analogous to (37) which still exhibit Principle B effects. Consider, then, the following scenario. In the year 2525 the US finally has a multi-party system. Unfortunately, all the candidates for president and vice-president are males that year. (I invoke the multi-party system simply to make each presidential candidate felicitous; and I invoke the male-only candidate scenario so that the presence of he is not odd.) Enter the moderator of a debate involving each presidential candidate and his running mate:

(41) Moderator of the debate:
    This - our second debate in our series - is the attack debate. The rules of this debate are very strict. You are allowed to attack your opponents all you want. But you’re not allowed to do any praising—neither of yourself, nor of your running mate, until you’re told that you’re allowed to. Moreover I—and only I—know when each of you will be told that you’re now allowed to go into praising mode.

Needless to say, this puts quite a stress on the candidates but

(42) *Each presidential candidate hoped that at least HE would soon be told that he could now praise him.
    (That is, he₁ is not sufficiently optimistic as to hope that his running mate will be informed of when he (the running mate) will be allowed to praise him₁.)

The impossible (or at least quite difficult) reading—but which should be quite salient—is exactly parallel to the bus accident scenario. The first HE is bound, the second he is bound and “sloppy” (the contrast set is another relevant individual—and in the reading at issue this is the candidate’s running mate), but the last him is strict. Yet here this reading is at best quite difficult. I think it might be slightly better than the Principle B violating reading in (15)—and this is not surprising given the mitigating influence of focus—but it still remains degraded (as other informants have confirmed).

So the bottom line is that a configurational constraint on co-indexing—like (18)—fails not only to account for the coreference-without-binding cases but also fails to account for cases in which the subject and object are not
co-indexed, but the binding pattern is such that they are in what I call the
"co-bound" configuration. Here too the relevant reading is blocked.

6.4.2 Some Proposed Solutions

6.4.2.1 Proposed Solution 1: Grice  What we see above is that a simple con­
straint against co-indexation really does not do the trick: it rules out certain
structures but it does not rule out the actual meanings (or interpretations) that
we want. There has, however, been a very seductive alternative in the literature
since at least as early as Dowty (1980) (see also Reinhart 1983; Sadock 1983, and
many since). This is to posit that the grammar contains nothing analogous
to Principle B: the effect is purely Gricean. The offending configurations are,
after all, just those in which a reflexive could have been used. Since a pronoun
is ambiguous and a reflexive is not, the use of the reflexive is more informative.
Hence by Gricean principles, if a speaker did not use a reflexive the listener will
assume that non-coreference (or non-binding) was intended. This, a single
account, covers all of the above cases, and removes the entire domain from
that of grammar.

This kind of solution is certainly appealing. It covers all of the above cases,
and it explains why, for example, Principle B violations such as (43) are not as
bad as those in (15)–(17) (this was pointed out to me by Jerry Sadock):

(43)  How do you manage to get everyone to vote for you?
    ?*Oh, I just praise me every chance I get.

(The logic here is that using myself is no more informative than using me.)
Notice that if a Gricean account could work, then direct compositionality
and variable-free semantics would be perfectly happy. The Gricean account
makes no use of a configurational constraint, there is no use of indices, and
the account would also extend to explain the badness of (19) without having
to posit additional structure around the post-copular constituent. Thus if a
Gricean account could be made to work, this paper could end here.

Unfortunately, a purely pragmatic account does not seem feasible, for the
effect resists any kind of cancellation and simply does not behave like other
known phenomena with pragmatic explanations. Note first that the Gricean
story is somewhat odd to begin with. It is not quite true that the meaning of
Bush saw him is less informative than the meaning of Bush saw himself. Under
the standard story, the meaning of the first is a non-constant function from
assignment functions to propositions while the latter is a constant function
and so they do not form a scale in the usual way found for the case of scalar
implicatures. (Similar remarks hold with respect to their meanings under the
variable-free view; there they do not even have the same type of meaning.) The
correct way to think of this is that the proposition denoted by *Bush saw him* is not fully specified, whereas *Bush saw himself* is. But other analogous cases—in which a sentence is either ambiguous or unspecified in some way—simply do not give rise to corresponding implicatures. Thus (44a) obviously does not implicate (44b) (even though (44c) is “more informative”):

(44)  
(a) *Bush’s mother likes him.*  
(b) *Bush’s mother likes someone (and that someone is someone other than Bush).*  
(c) *Bush’s mother likes Bush.*

Moreover, analogous cases where we have an ambiguity or underspecification are perfectly happy when followed by extra clarificatory material:

(45)  
(a) *Barbara praised him... that is to say, George.*  
(b) *Barbara praised him—that is to say, she praised George.*

(46)  
(a) *I went to the bank—that is to say, the river bank.*  
(b) *I went to the bank—that is to say, I went to the river bank.*

Yet such clarification does nothing to remove a Principle B violation:

(47)  
(a) *Bush praised him—that is to say, himself.*  
(b) *Bush praised him—that is to say, he praised himself.*

Sentences (47a, 47b) do not even remotely smack of an improvement, and indeed I can think of no way at all to cancel the “implicature” or remove the ambiguity. If this is a Gricean phenomenon, it behaves unlike others that we know of, and unless and until we have an explanation for this it appears that we probably have to live with the conclusion that the effect is located in the grammar. (Again, however, if it is not, then DC will be perfectly happy.)

6.4.2.2 Proposed Solution 2: Principle B Plus a Transderivational Principle  
We now arrive at the solution most commonly accepted in at least the modern formal semantics (LF-based, non-DC-based) literature. This is to posit that there really are two principles at work here. One is what we can call “core Principle B” and is the principle given in (18): it is a principle regulating the distribution of indices. It will block the relevant reading for (17b) (in which the pronoun has to be bound), and it will block the representation of (15b) shown in (20a). As to the fact that this is not enough to block the relevant understanding of (20b) (where the pronoun can remain free) and (28) (where there is co-binding), the idea is that there is a second “transderivational” principle—which blocks these readings in virtue of the fact that—were it not for Principle
B—the bound reading would have resulted in the same interpretation. This has been stated in a couple of different ways; let me take the formulation in Büring (2005) which is closely based on the formulations in Fox (2000) which in turn is based on Heim (1993) which in turn is based on Grodzinsky and Reinhart (1993):

(48) For any two NPs $a$ and $\beta$, if $a$ could bind (i.e. if it c-commands $\beta$ and $\beta$ is not bound in $a$’s c-command domain already), $a$ must bind $\beta$, unless that changes the interpretation (use of boldfaced mine, PJ)

This is intended to account for both of the cases discussed above in §6.4.1. Consider first the LF in (28)—where neither pronoun binds the other. Example (48) will block this LF, because the interpretation of this LF is identical to one in which the subject pronoun binds the other. Consider the LF in (20b)—where the pronoun is “free”—but where there is an “interpretation” where it happens to pick up Bush. (Technically this is not quite the right way to put it: the sentence is a non-constant function from assignment functions, and we apply it to one where $7$ is assigned to Bush. But I will continue to use the more informal terminology here.) The idea of (48) is that that LF cannot have that interpretation. The reason is that there is another LF (i.e. (20a)) in which the pronoun is bound and which results in the same interpretation. Note that “interpretation” here does not mean just the output of the compositional semantics—in order for this to block the reading where the free pronoun in (20b) “picks up” Bush, “interpretation” has to mean the output of the processing system.

It seems to me that the claim that there is a principle like (48) cannot survive serious scrutiny; it has problems which seem insurmountable.\(^\text{10}\) Perhaps the most important of these derives from the fact that “could bind” in this case has to overlook “Core Principle B” violations! That is, Principle B (as given in (18)) says that in (15b), the subject cannot “bind” him. (It rules out co-indexation and therefore binding.) But the non-co-indexation reading will have to be ruled out in virtue of the fact that the co-indexation (binding) structure would have been possible and would have yielded the same interpretation. And yet co-indexation would not have been possible. Thus we need to rule out one case in virtue of another “bad” case. This fact alone, it seems to me, should be enough to make us reject this right away. Note that competition effects—where one representation/interpretation pair is blocked

\(^{10}\) To be fair, (48) is designed not only to account for the full range of Principle B effects, but also for a puzzle surrounding ellipsis due originally to Dahl (1974). The reader should consult Fox (2000) for a full discussion. I have no speculation to offer here as to how to account for Dahl’s puzzle, but I hope the remarks below convince the reader that (48) is quite implausible, despite this apparent side benefit.
in virtue of a competing representation/interpretation pair—may well exist as constraints on processing. For example, Gricean principles themselves rely on speaker/hearer awareness of competition effects. But this is a case where one pair is blocked on the basis of a competing bad representation/interpretation pair—a situation which, as far as I know, is unprecedented.

In fact, it is worth taking very seriously what kinds of competition effects are known to exist in processing. One kind are competition effects based on form—these include typical Gricean effects. A listener hears form $F_1$, which is vague (or perhaps ambiguous) and can be literally true in a certain set of situations $X$ as well as in a broader or different set of situations $Y$. However, there is another (generally equally easy to say) form $F_2$ which can only be true in $X$, and hence $F_2$ makes a stronger (or less ambiguous) statement than $F_1$. The listener infers that had the speaker intended to convey that the set of situations $X$ holds, s/he would have said $F_2$, so it must be the case that we are not in situation $X$ (i.e. that the proposition expressed by $F_2$ is false). So, here is a case where one form blocks a particular interpretation for another form. The other kind of competition effects that are well-known concern the case of competing meanings. Here there is a single form with more than one possible interpretation, but because one meaning $M_1$ is (for whatever reason) more accessible than another meaning $M_2$, $M_1$ blocks $M_2$. (Garden-path effects are examples of this phenomenon.)

But the kind of competition here is neither of these (and we have already seen that trying to reinterpret this as a Gricean effect unfortunately does not seem viable). For here the two forms are the same, and the two meanings (or, interpretations) are the same! The claim here is that one way to derive a certain interpretation for a given form is blocked in view of the fact that there was another way to derive the same interpretation from the same form. (And of course, as has already been pointed out, the "other way" is itself actually bad.) I know of no processing principles which would motivate this kind of competition, and know of no plausible story which has ever been proposed to explain this. Note too that this cannot be a direction or strategy to the processor specifically for how to assign meanings to "free" pronouns since it must also block the cobound cases. But neither can it be a constraint on LF representations (which do not have a pronoun bound less locally if it can be bound more locally without changing the interpretation) since this will not constrain the computation of free pronouns (which is not given by the compositional semantics).\footnote{At first glance it also appears that the existence of a constraint such as (i) would undermine the standard story regarding the possibility of both strict and sloppy readings in:

(i) John, loves his; mother and Bill does too.}
6.5 A Direct Compositional, Variable-Free Analysis

6.5.1 The Analysis

What we have seen is that configurationally based accounts of Principle B effects (based also on co-indexing) seem extremely unlikely to be correct; the complexities involved should make us very suspicious that such accounts are really missing the point. Can we do better with a direct compositional approach—one which skips the representational steps and tries to capture the effect in terms of the syntactic and semantic combinatorics? I will propose that we can.

My proposal is embedded within a Categorial Grammar syntax and makes crucial use of one fundamental property of CG: each syntactic category is an encoding of distributional properties (it also encodes semantic types). With respect to verbs (and other “function” or “complement-taking” categories), the category label itself encodes exactly what complement an item can take. Thus an expression of category $A/RB$ takes a $B$-category expression to its right to give a resulting expression of category $A$, and an $A/LB$ takes a $B$ to its left to give an $A$. (As noted in §6.2, I assume that there are also expressions which take their arguments as infixes and expressions which themselves infix into their arguments. For details, see Jacobson (1992).)

As noted earlier, an expression with the general distribution of $A$-type expressions but which contains within it an “unbound” pronoun of category $C$ is an expression of category $A^C$. This means that material with a pronoun within it is not quite the same category as corresponding material without the pronoun in it: and so if we have a verb with, for example, the category $(S/NP)/NP$ (an ordinary transitive verb) it cannot combine with an NP that contains an unbound pronoun within it. But there are two unary rules which shift it into something which can. One is the $g$ rule (which shifts it to $(S/NP)NP/NPNP$—the semantic effect here is to allow the verb to take a function of type $<e,e>$ as argument—and it passes up the argument slot of that function for later binding. (In other words, the result after the verb combines with its object is a function of type $<e,<e,t> >$ rather than an $<e,t>$. The second rule is the $z$ rule, which shifts an ordinary transitive verb into one of category $(S/NP)/NP^NPP$. Here the semantic effect is again that a function of type $<e,e>$ is expected as object argument, but here the argument

The apparent problem is that—in the way it is usually put—the strict reading arises when the first his is a free pronoun not co-indexed with John. But (i) would then disallow such a representation to be paired with an interpretation where his is free and happens to pick out John. This has, in fact, worried past researchers (see Reinhart 2000 and Büring 2005 for discussion). Actually, though, this is not really a problem, for one could allow only the “bound” representation for his while still allowing the strict reading, by using the strategy outlined on p. 210 above.
slot of that function is "merged" with the subject slot. Thus after the verb takes its object the result is of type \(<e,t>\); we can say that in this case the pronoun within the object is "merged with" or "bound to" the subject slot.

In my previous work, I treated pronouns themselves and material containing (unbound) pronouns as having exactly the same syntactic category (and semantic type). Both were of category \(\text{NP}^{\text{NP}}\) (and both denote functions of type \(<e,e>\)). But what we learn from a close look at Principle B effects is that bare pronouns actually do not seem to have the same distribution as pronouns sunk within further material. (I return below to the case of pronouns within argument-type PPs as in *He; gave a book to him;.) Let us suppose, then, that bare pronouns are listed in the lexicon not as just \(\text{NP}^{\text{NP}}\) but actually they contain a special feature \([p]\), where \(\text{NP}^{\text{NP}}\) and \(\text{NP}[p]^{\text{NP}}\) are not the same thing. All of this is just to say: pronouns cannot occur except where they are specifically sanctioned. (I return below to the fact that they always seem happy in subject position.)

The informal idea, then, is as follows. Take an ordinary transitive verb like \(\text{praise}\), and assume that it is listed in the lexicon with syntactic category \((S/\text{NP})/\text{NP}\) and with meaning of type \(<e,<e,t>>\). Moreover, its meaning is just as we have always thought: it characterizes the set of all ordered pairs of praisers-praisees, including the reflexive pairs. But because of its syntax, this verb cannot directly take a pronoun in object position. A pronoun has a special category and it cannot occupy ordinary NP argument slots. Note of course that the lexical item \(\text{praise}\) could in any case not take a pronoun in object position—it would have to be mapped first into \(z(\text{praise})\) or into \(g(\text{praise})\). But the new point here is that even after \(\text{praise}\) undergoes one of the other rules it is still unable to take an ordinary pronoun in object position. \(g(\text{praise})\) is of category \((S/\text{NP})^{\text{NP}}/\text{NP}^{\text{NP}}\) and \(z(\text{praise})\) is of category \((S/\text{NP})/\text{NP}^{\text{NP}}\). So neither would be happy with a bare pronoun.

We thus posit one more unary rule—a rule that allows the lexical item \(\text{praise}\) to map into a new verb which (once it undergoes one of the two rules above) will be able to take a bare pronoun in object position. Syntactically this rule maps the NP argument slot in object position into an NP\([p]\) argument slot. The interesting part is the semantics: the new verb becomes undefined for the reflexive pairs in the original denotation. Thus the rule is as follows (it would need to be generalized for the case of three-place verbs, but this is straightforward and is left as an exercise for the reader):

\[\text{(49) Let } a \text{ be an expression of the form }<[a]; (S/\text{NP})/\text{NP}; a'>. \text{ Then there is an expression } \beta \text{ of the form: }<[a]; (S/\text{NP})/\text{NP}[p]; \text{irr}(a') >\text{ where for any function } f \text{ of type } <e,<e,t>>, \text{irr}(f) \text{ is the function}\]

mapping any \( x \) and \( y \) as follows:

\[
\text{irr}(f)(x)(y) = 1 \text{ if } f(x)(y) = 1 \text{ and } x \neq y
\]

\[
\text{irr}(f)(x)(y) = 0 \text{ if } f(x)(y) = 0 \text{ and } x \neq y
\]

\[
\text{irr}(f)(x)(y) \text{ is undefined if } x = y
\]

The happy part of this is that Principle B effects derive from the syntactic and semantic combinatorics—the account directly regulates a correspondence between what syntactic categories can occur where (in this case, it regulates the distribution of pronouns) and meaning. The even happier part is that this—with no further ado—accounts for all three of the cases that we have seen above. Let us take them each in turn.

### 6.5.2 Accounting for all Three Principle B Cases

#### 6.5.2.1 The Bound Case

As noted above, the mapping of ordinary praise to \( \text{irr} \) (praise) is still not enough to allow a bare pronoun in object position: this becomes possible only with the additional application of \( z \) or \( g \). Consider the badness of the “bound” case in (50):

(50) *Every candidate; praised himi.*

Here the pronoun is “trying” to be bound by the subject slot, and so a sequence of steps which produced this reading would involve the application of \( z \) to \( \text{irr} \) (praise), whose category and meaning are as follows:

(51) \( z(\text{praise}_{\text{irr}}); (S/NP)/NP[p]; \lambda f[\lambda z[\lambda x[\lambda y \neq x \text{ praise}'(x)(y)][f(z)](z)]] = \lambda f[\lambda z_{\neq f(z)} \text{ praise}'(f(z))(z)] \)

There is nothing wrong with this meaning (indeed we will see momentarily that it is useful), but nonsense ensues when this takes as argument the ordinary meaning of a pronoun: that is the identity function. This is what happens in (50). Thus the meaning of the VP in (50) is the function characterizing the set of self-praisers, but defined only for those individuals who are not equal to themselves (hence of course undefined for everyone).

As noted directly above, there is nothing incoherent about the meaning of \( z(\text{irr}(\text{praise})) \) (syntactically, this verb will require a bare pronoun in object position\(^{12} \)); the incoherence ensues only when (51) is applied to the identity function over individuals. This means that if a bare pronoun could have some other meaning—in addition to the identity function on individuals—then it could combine with this verb. And in fact, in Jacobson (2000) I argued that paycheck pronouns are derived from ordinary pronouns via the Geach rule; a

\(^{12}\) This assumes that non-pronouns cannot appear in slots marked NP[p], just as items of category NP[p] cannot appear in ordinary NP slots.
paycheck pronoun is thus the identity function over functions of type \langle e, e \rangle. What this means is that we should be able to combine \( z(\text{irr}(\text{praise})) \) with the paycheck meaning of a pronoun, thereby binding "into" the argument slot of the paycheck. This prediction is correct: there is no problem with the reading of (52) notated above, which involves this verb:

(52) Every Democratic candidate, tried to distance himself from his father, while every Republican candidate continually praised him.

Again I use indices in the obvious way merely to indicate the intended reading.

6.5.2.2 The Free Case

Next consider the case of a "free" pronoun, as in (53).

(53) *Bush praises him.

The difficulty with the standard account is that there was no non-tortured way to rule out an understanding of this sentence in which \( him \) remains free and just happens to pick up Bush. But here this is ruled out. A "free" pronoun in the account here is one which was introduced via the application of \( g \) (or, really, applications of \( g \) all the way through the semantic combinatorics)—which has the effect of allowing a function of type \langle e, e \rangle to occur in some normal e-slot, and "passes up" the binding. But in a free case, we never get an application of \( z \)—and so the final result is a function from individuals to propositions. This, then, is one possible meaning for a sentence such as Bush praises him. It is, however, a function which is defined only for individuals other than Bush.

To explicitly show the details, I need to point out that the free reading of the derivation will involve treating the subject as a generalized quantifier (i.e. the "lifted" meaning of Bush). With that, we can spell out the full derivation of Bush praises him:

\[
\begin{align*}
(54) & \quad \text{praise; } (S/NP)/NP; \text{ praise'} \\
& \rightarrow \text{irr praise; } (S/NP)/NP[p]; \lambda x[\lambda y_{y \neq x}[\text{praise'}(x)(y)]] \rightarrow g \text{ praise; } (S/NP)^{NP}/NP[p]^{NP}; \lambda f[\lambda z[\lambda x[\lambda y_{y \neq x}[\text{praise'}(x)(y)](f(z))]]] \\
& = \lambda f[\lambda z[\lambda y_{y \neq f(z)}[\text{praise'}(f(z))(y)]](\lambda x[x])] \\
& = \lambda z[\lambda y_{y \neq z}[\text{praise'}(z)(y)]] \\
& \quad \text{(thus note that } g(\text{irr}(\text{praise'}))(\text{him'}) = g(\text{irr}(\text{praise'})) \\
& \text{Bush; } S/(S/NP); \lambda P[b]) \rightarrow g \text{ S}^{NP}/(S/NP)^{NP}; \\
& \lambda R_{<e, <et>}[\lambda x[\lambda P[P(b)][R(x)]]] = \lambda R[\lambda x[R(x)(b)]] \\
& \text{Bush praises him; } S^{NP}; \lambda R[\lambda x[R(x)(b)]](\lambda z[\lambda y_{y \neq z}[\text{praise'}(z)(y)]])(x)(b)) = \\
& \lambda x[\lambda z[\lambda y_{y \neq z}[\text{praise'}(z)(y)](x)(b)] = \\
& \lambda x[\lambda y_{y \neq x}[\text{praise'}(x)(y)](b)] = \lambda x_{x \neq b}[\text{praise'}(x)(b)]
\end{align*}
\]
(The last line in the steps which are notational simplifications of the meaning is not an ordinary \(\lambda\)-conversion, but the reader can verify that it is the correct equivalence.) In the end, then, the sentence denotes a function from individuals to propositions, but one defined only for those individuals distinct from Bush. Hence the reading indicated in (53) amounts to a presupposition violation.

6.5.2.3 The Co-Bound Case Now let us consider how this accounts for the impossibility of (55):

(55) *Every candidate; \(i\) thinks that \(he; i\) said that \(he; i\) praised him; \(i\).

Recall that the problem for the traditional account is that this cannot be blocked by a simple constraint on co-indexing since the two most deeply embedded pronouns could have different indices, but where the subject of say "binds" the subject of praise and every candidate "binds" the subject of say and the object of praise. (I of course use this terminology loosely since we have already seen that there is no obvious sense in which one NP "binds" another, but I use this as loose terminology for the LF shown in (28).) It will be useful at this point to give an informal rendering of this LF by means of the diagram shown in (56):

(56) \[
\text{every candidate} \quad \xrightarrow{\text{thinks}} \quad \text{he say he praise him}
\]

The question for the variable-free view, then, is to determine whether or not there is any kind of analogous problem. Before answering this, we first need to consider what happens in general when there are two pronouns which are "bound" by the same thing. Consider, for example, (57):

(57) Every man; \(i\) in our dog park thinks that the woman he; \(i\)'s engaged to should walk his; \(i\) dog.

As discussed in Jacobson (1999), the two pronouns in (57) necessarily are not semantically "linked" in terms of the meaning of the embedded S. In the normal view, they can correspond to the same variable—but there is nothing analogous to this here. Rather, the meaning of the S embedded under think is the two-place relation

(58) \(\lambda x[\lambda y[\text{the woman that } x \text{ is engaged to should walk } y \text{ 's dog}]]\)
How then do we get the reading in (57)? In this case this comes about by two applications of \( z \) on \( \text{thinks} \)—which has the effect of "merging" these two slots (each is "merged" with the subject slot). The details are space-consuming and so the interested reader can consult the full details in Jacobson (1999).

Let us consider the LF (28) as a way in the standard theory to get the meaning for (56) indicated above. There is nothing exactly analogous to this in the variable-free view, but the closest analog would be a derivation in which \( \text{say} \) undergoes both \( z \) and \( g \) in such a way that \( z \) "merges" its subject slot with the argument slot created by the object pronoun, and \( g \) "passes" up the other pronoun (the subject of \( \text{praise} \)) for higher binding. \( \text{Thinks} \) then undergoes two applications of \( z \)—one "binds" the subject position of \( \text{say} \) to the subject position of \( \text{thinks} \); and the other "binds" the passed-up pronoun slot (the object of \( \text{praise} \)) to the subject position of \( \text{thinks} \) which then merges these two. It is hoped that the diagram in (59) will help illustrate this, and will clarify the parallel between this derivation and the LF in (28) (I omit a number of applications of \( g \) from this picture):

\[
(59)
\]

But note that \( \text{praise} \) is the irreflexive version of \( [[\text{praise}]] \)—as it must be in order to have a pronoun in the object position. It follows from this that \( \text{he praise him} \) is itself \( [[\text{praise}_{irr}]] \). And given this, the derivation sketched above gives nonsense at the end of the day, as we would expect. To show this out formally we need to point out that undefinedness (like other presuppositions) projects upwards as, for example, the case of the gender presupposition in (60):

\[
(60)
*\text{John said that every man; loves her; mother.}
\]

Given this, consider the full details of the derivation sketched above.

\[
(61)
\text{z(g(say))} = \lambda R[\lambda w[\lambda z [ z \text{ say } R(w)(z) ]]]
\]

said he praised him:

\[
\begin{align*}
&= \lambda R[\lambda w[\lambda z [ z \text{ say } R(w)(z) ]]] (\lambda x[\lambda y_{\neq x}[\text{praise}' (x)(y)]]) \\
&= \lambda w[\lambda z [ z \text{ say } \lambda x[\lambda y_{\neq x}[\text{praise}' (x)(y)]] (w)(z)]] \\
&= \lambda w[\lambda z [ z \text{ say } \lambda y_{\neq w}[ \text{praise}' (w)(y)] (z)]] \\
&= \lambda w[\lambda z_{\neq w}[ z \text{ say } [\text{praise}' (w)(z)]]
\end{align*}
\]
\[ z(\text{think}) \cdot x \mapsto z(x) \cdot R(x)(x) \cdot x \mapsto \lambda x \cdot x \mapsto \lambda x [x \text{ thinks } R(x)(x)] : \]
apply to above: \( \text{thinks he said he praised him} = \lambda x \cdot x \mapsto \lambda x \cdot x \mapsto [x \text{ thinks } x \text{ said } x \text{ praised } x] \)

Thus, exactly as in the case of (50), the VP here is nonsense and is not defined for any individual.

Of course it should be pointed out that there are other ways one could attempt to get the relevant reading; in order to show that this is “nonsense” under all such derivations one would need to show out every possible derivation. Space precludes doing this here; hopefully the above gives a flavor of why, once we are dealing with the irreflexive version of praise, we will have no way to get this kind of a Principle B violation.

6.5.3 Principle B Connectivity Effects

To complete the discussion, note now that the appearance of Principle B effects in a copular case such as (62b) is completely expected without positing any complex analysis regarding the representation of the post-copular constituent:

62. (a) What every candidate \( i \) does (every chance he gets) is praise his \( i \) mother.
(b) *What every candidate \( i \) does (every chance he gets) is praise him \( i \).

The “null hypothesis” is that the post-copular constituent has no deleted or silent material surrounding it, and the appearance of Principle B effects does not require us to abandon this hypothesis, since the effect is encoded purely locally within the post-copular VP, and so there is no need to surround it by additional material in order to satisfy a representational, non-strictly local constraint such as the one in (18). Roughly, then, the compositional semantics for (62a) is as represented in (63a), and the compositional semantics for (62b) is shown in (63b):

63. (a) the property \( P \) such that \( P \) is true for every candidate is the property of \( z \)-praising the mother function
(b) the property \( P \) such that \( P \) is true for every candidate is the property of \( z \)-praising (irr) the identity function
   (i.e., the property of being a self-praiser, defined only for those individuals not equal to themselves)

For a case like (19a) there are two ways one might try to get the impossible reading:

19. (a) *What Bush \( i \) always does (on the campaign trail) is praise him \( i \).
One would involve application of z to irreflexive praise in the post-copular constituent, and this will yield nonsense exactly as in the case of (63b). The other is to let him be “free”—by which is meant that g will apply to irreflexive praise. In the end, then, the entire sentence will denote a function from individuals x to the proposition that “what Bush does is praise x”, but where this function is defined only for individuals not equal to the subject of praise. Thus the function is undefined for Bush, in much the same way as in the ordinary case of (15b). (The full details have to do with the entailments of a copular sentence; but it is clear that under any way to think of the semantics of the copula here it will follow that him cannot be Bush. We will leave the full details for the interested reader to supply.)

6.6 Open Questions

There are many well-known complexities surrounding Principle B effects, and some of these raise questions for this analysis as well as for more “standard” analyses. There are also some questions raised by this analysis in particular. While there is no way I can cover all of the questions that arise, this section will conclude with a look at some of these.

6.6.1 Some Mechanical Issues

First, the analysis is not complete and some of the details have not been fully spelled out above. The analysis needs to be extended to the case of three-place verbs (as in (64)); it needs to be extended to cover the case of a pronoun within an argument PP (as in both (64) and (65)) and it needs to be refined to account for the fact that pronouns are always happy in subject position:

(64)  (a) *I introduced every man, Bushi to him.
        (b) *Bushi/every man, introduced me to him.

(65)  *Bushi/Every man, talked about him.

Without giving the full details here, I believe that most of the requisite machinery is fairly straightforward. For the case of three-place verbs, we simply assume that there is a generalization of the irr operation (such extensions are needed in general for three-place verbs), which allows the feature [p] to occur on non-subject argument slot, and removes from the denotation of the verb all triplets which are such that the same individual occurs in the relevant argument slot and in one of the higher slots. (Recall that I am assuming a Wrap analysis for (64) in which every man/Bush is introduced later than the PP and is “wrapped” in (infixed) during the syntactic composition.)
The fact that a pronoun cannot be embedded in an argument PP is hopefully easy to capture with judicious use of "feature passing". More precisely, prepositions like to or about which yield argument are of category PP/NP and can map into PP[p]/NP[p]. Hopefully the fact that the [p] feature is passed from argument to result in this case follows from something more general, but I will not pursue this here. I assume that the meaning of such prepositions is the identity function, so the meaning of to him will be the same as the meaning of him (it is the identity function on individuals). (We further assume that the convention that NP[p] can occur only where specifically asked for holds equally well for PP[p].) As to the fact that pronouns are always happy in subject position, we might assume that the feature [p] which we are concerned with here occurs only on accusative pronouns. None of these suggestions are terribly insightful, but they would seem to be straightforward mechanical extensions of the type needed for other phenomena.

6.6.2 Lexical Variation in the Strength of Principle B Effects

A more interesting wrinkle concerning Principle B effects which, I believe, has not been sufficiently appreciated in past literature is that the strength of the effect is very lexically specific. Some verbs allow Principle B violations (with a bit of work) relatively easily; others resist these. This observation, by the way, will account for a glaring contradiction in the literature. Fiengo and May (1994) claim that Principle B effects survive ellipsis, and use an example along the lines of the following:

(66)  
(a)  *John introduced heri to everyone before shei did.
(b)  *John introduced heri to Bill, but shei already had.

(All of these improve somewhat if the subject is she herself rather than just she, and this improvement holds throughout all of the sentences cited below. I will not speculate here on why this is so; I will however keep the data consistent by never using this form.) But on the other side of the coin it is well-known that focus ameliorates Principle B effects, and cases such as (67) are typical of the cases cited in the literature to show this:13

(67)  
(a)  Mary pities himi, Sue pities himi, and ??I even think that HEi pities himi.
(b)  ??I only think that HEi pities himi.

13 The literature varies on whether or not these are perfect or whether they are just improved with focus on the subject but still not perfect. My very unsystematic check with (a very small sample of) informants would seem to indicate the latter—I have consulted with three speakers (obviously not statistically significant)—and they find these to be still somewhat off. Obviously this needs to be checked further, but I will assume here and below that focus merely improves but does not completely remove the Principle B effect.
The fact that the effect is ameliorated by focus but not by ellipsis should come as a complete surprise (it is odd that this has not, to my knowledge, been commented on in the literature). The two often behave alike in matters of this type (in fact, ellipsis actually requires focus, so it is hard to understand why a difference occurs); and from what we know about the two we would expect that—if anything—ellipsis would be “less fussy” here.

But there really is no contradiction: the difference stems from the fact that pity is what we might call a relatively weak Principle B verb while introduce shows very strong Principle B effects which resist amelioration in any context. Thus note that (68) remains awful despite the fact that the subject is focused, and (69) shows that VP ellipsis ameliorates the violation with pity:

\[(68)\]
\[(a)\] *John introduced her\(_i\) to Bill, Sam introduced her\(_i\) to Bill, and I even think that SHE\(_i\) introduced her\(_i\) to Bill.

\[(b)\] *I only think that SHE\(_i\) introduced her\(_i\) to Bill.

\[(69)\]
\[(a)\] ??John pities her even more than she does.

\[(b)\] ??John pities her, and she does too.

In fact, the effects are so strong with introduce that one might be tempted by a suggestion of Tamina Stephenson (p.c.) that there are really just two separate verbs introduce. The first (call it introduce\(_1\) —which occurs in all of the examples above—is lexically irreflexive: one simply cannot introduce\(_1\) oneself to another. The other is introduce\(_2\) which is only reflexive and is therefore the verb that occurs with a reflexive as in Mary introduced herself to Bill. This seems tempting, but unfortunately it would make it quite a mystery as to why (70) is perfect:

\[(70)\]
\[(a)\] Mary introduced herself to Bill before Sam had a chance to.

\[(b)\] Mary introduced herself to Bill because Sam wouldn’t.

I should point out that the asymmetry between “Principle A” effects with introduce in (70) and Principle B effects is actually a bit of a mystery given certain plausible views of the semantic contribution of a reflexive, but it seems to me that any reasonable story about why (70) is good would make it quite difficult to maintain the two verbs introduce theory. I will assume, then, that some other explanation is needed for the different Principle B strengths with different verbs.

It seems quite likely that the explanation will ultimately lie in a suggestion by Heim (1993) that pronouns are of individual concepts (of type <s,e>) and that Principle B requires that the subject and object be distinct individual concepts. To use Heim’s informal terminology, apparent Principle B violations are possible when the subject and object are extensionally equivalent, but
characterize the same individual "in different guises". The extent to which a verb allows apparent violations will depend on the extent to which its meaning is such that it makes salient the possibility of the subject and object having the same extension while being distinct individual concepts.

6.6.3 Raising to Object Cases

Whenever I present this analysis, someone inevitably asks about "Raising to Object" (or "ECM") verbs, as in:

(71) *John, expects him, to win.

These, of course, present a problem for the analysis only if one believes that *him is a subject in the lower clause and hence not an argument of *expect. But in the general tradition assumed here, there is plenty of independent reason to believe that *him to win is not a clausal complement of *expect. (For example, this form of the verb *expect has—as noted in Rosenbaum (1967)—a passive counterpart. If Passive is a rule rearranging the argument structure (the syntactic and semantic argument slots) of a verb, then of course *him in (71) must be an argument of *expect.

Within the Categorial Grammar and related literature, there are two distinct ways which have been proposed to treat Raising to Object verbs, and either one will be compatible with the proposal here. In one (see especially Dowty 1985) *expect and *persuade have the same argument structure in the lexicon: they both take an infinitive VP and then an NP to give a VP. The familiar "raising" vs. "control" are simply a matter of fine-grained facts about the meanings of these verbs (and control itself is a consequence of entailments associated with the verbs). A somewhat different analysis is developed in Jacobson (1990) which more closely mimics the traditional "raising to object" analysis (Rosenbaum 1967). Here *expect is—as in the standard view—listed in the lexicon as taking an S complement, and it denotes a relation between individuals and propositions. However, it (obligatorily) function composes with its complement, and so *expect to win results, is an (S/NP)/NP, and the object "wraps" in. Note that this analysis is also perfectly compatible with the present account of Principle B effects: *expect to win is a complex-transitive verb; it cannot take a pronoun in the (wrapped-in) object position until it undergoes the reflexive rule. The difference between the two accounts lies simply in just how it is that the phrase *expect to win is composed, but the end result is the same. In other words, as long as *him is an argument of *expect to win we have no problem here.14

14 In comments on this paper at the Direct Compositionality Workshop, Danny Fox claimed that there seems to be a difference between "ECM" verbs and others in their Principle B effects, in that in the
6.6.4 Amelioration under Focus

There remains the question as to why focus ameliorates the effect. An answer to this question is of particular urgency for the account here since the amelioration has often been taken to support the existence of something like the transderivational constraint in (48). For in certain cases when the subject is focused, the representation in which the object is “bound” by the subject actually has different truth conditions from the non-coindexed structure (interpreted with “coreference”), and so (48) will not apply. There are, however, reasons to doubt this explanation. First, the truth conditions are indeed changed if the focused subject falls within the scope of only, but they are not changed with simple focus (or when the subject falls within the scope of even) yet still there is an amelioration (see, e.g., (67a)). Second, the amelioration does not seem in general to be complete. The facts seem quite fuzzy (as a few other informants have confirmed) whereas we would expect a complete disappearance of the effect (with only) if the above story were correct.

One hypothesis is that the sentences with focus on the subject continue to exhibit a presupposition violation, but since focus makes salient the alternatives (in which there is no such violation), we are able to accommodate this violation. The fact that focus ameliorates the situation but does not make it perfect is consistent with the hypothesis that accommodation is at work here. This explanation, however, cannot be the whole story—since it gives no insight into why the violation persists so strongly in the case of introduce. (Notice that the more standard explanation given above suffers from the same problem.) I thus leave this as an unsolved problem.

case of the former, there is no survival of this effect under ellipsis; thus Fox contrasts cases like (i) with cases like (ii) (Fox’s actual sentences involve a proper name in object position in the first clause which introduces an irrelevant complication into the discussion so I have changed these; the distribution of *s here is essentially that reported by Fox (modulo the fact that his sentences are slightly different as noted above)):  

(i) Mary likes him. *He does too. <like himself>
(ii) I expected him to win the race. He did, too. <expect himself to win the race>

But we have already seen that Principle B effects in general can be ameliorated by ellipsis and I think that the contrast between (i) and (ii) is anything but sharp. We see above, moreover, that the strength of Principle B effects is subject to a lot of lexical variation, and expect might just be one of the weaker Principle B verbs. Other Raising to Object verbs seem to me to show the effect somewhat more strongly:

(iii) I consider him to be competent. ??He does too.

I thus do not think this is a systematic difference between Raising to Object verbs and other verbs.
6.6.5 Conjoined Objects

Anna Szabolcsi (p.c.) has raised the following objection to this analysis. Consider (72) (note that the pronoun here is not intended to be understood as "coreferential" with the subject of *likes*):

(72) John₁ thinks that Bill₁ likes (both) himself₁ and him₁.

Szabolcsi's worry goes as follows. Since we have the pronoun *him* in object position, then we must be dealing with *likes*ₐᵣₑ—the verb which characterizes the set of pairs \((x, y)\) such that \(x\) likes \(y\) and \(x \neq y\). Actually this is a slightly oversimplified way of stating the problem, for here we have a plural object. The worry, however, is that the meaning of *likes* which takes this plural object would have to be such that it takes a group \(X\) and distributes such that for each atomic part \(x\) of \(X\), the subject stands in the *likes*ₐᵣₑ to \(x\). But then how could we possibly have the reflexive here? (Note that the problem is independent of one's formal account of reflexives.)

But there is actually no problem, once we note that the NP *himself and him* is itself not a pronoun—and hence presumably does not carry the feature \([p]\). Hence *likes* is not the irreflexive variant (or, more precisely, is not the variant which would distribute to the irreflexive *likes*). In other words, the \([p]\) feature is not "passed" from one conjunct to the entire coordinated constituent (there is no reason to think it would be; it is passed only in the case of NPs within PPs). This hypothesis is easily tested: we need merely to see whether Principle B effects in general survive coordination. It seems to me (and the three other informants whom I have consulted) that they do not:

(73) (a) Bush₁/every man₁ praised (both) Barbara and him₁.
(b) Bush₁/every man₁ praised (both) him₁ and Barbara.

This is not the end of the story, for Szabolcsi (p.c.) informs me that the Hungarian situation is much like the English situation in all other respects, but that Principle B effects do survive in cases analogous to (73). (Yet cases analogous to (72) are fine.) Obviously, then, the Hungarian case needs to be looked at in greater detail.

6.6.6 The "Except Bill" Case

Recall the discussion on pp. 211–12 concerning sentences like (23):

(23) Bill₁ thinks that no one at the moment is living in his₁ house.

The question raised in that section was whether or not one could adopt a constraint on assignment functions such that for any two distinct variables,
their value on any assignment function $g$ must be distinct individuals. As discussed there, this would mean that (23) says that Bill thinks that no one except himself is living in his house, which is clearly the wrong meaning.

Interestingly, the proposal here suffers from a version of this same problem—but in this case the "problem" arises only in the contexts where we find Principle B effects. Thus (23) is unproblematic (Bill himself is not exempted), but we do get the "except for Bill" reading in local cases like (74); I leave it to the interested reader to verify why we get this prediction:

(74)  
(a) Bill$_i$ thinks that no one voted for him$_i$.
(b) Bill$_i$ thinks that no one likes him$_i$.

In this case, though, it is not clear to me that there is anything pernicious about this prediction. The literature seems to assume that here no one literally should not exclude Bill. Yet I think our intuitions are not at all clear on this. The prevailing wisdom is to the extent that it "feels like" no one does not include Bill. This is for pragmatic reasons—literally the subject is interpreted to be no one including Bill. This is a perfectly reasonable conclusion, but it seems to me to be equally reasonable to assume that the actual literal meaning here does exclude Bill. It is certainly clear that the "no one except Bill" understanding is much more robust in (74) than in (23) (it seems entirely absent in (23)).

6.6.7 Full NPs in Object Position?

So far, I have taken the Principle B effect to be relevant only to pronouns in object position, as in (75a). But we see (roughly) similar "non-coreference" effects when proper names and other full NPs are in object position, as in (75b) or (75c)

(75)  
(a) *Bush$_i$ praises him$_i$.
(b) *Bush$_i$ praises Bush$_i$.
(c) *He$_i$ praises Bush$_i$.

The account developed above says nothing about cases such as (75b), and I think that this is probably a good thing. The situations under which names (or full NPs in general) can be coreferential to other names (or pronouns) are very poorly understood. But in any case whatever accounts in general for the appearance of "Principle C" effects will automatically extend to account for (75b) and (75c) and so I have deliberately confined my account of Principle B effects to be relevant only to the appearance of pronouns in object position.

Suppose, however, that we do find evidence that (75b) and (75c) go beyond the routine "Principle C" effect and violate some additional principle. Should
that turn out to be the case, the analysis here can be extended: we merely say
that a verb like praise maps into an irreflexive verb if it takes anything but an
NP with a reflexive feature on it in object position. For now, however, I will
assume that irreflexive praise requires only a non-pronominal object.

In the remarks above I appealed to "Principle C" effects to account for (75a)
and (75b). Of course it remains to spell out just what these are. Note that the
relevant principle is another apparent counterexample to the hypothesis of
direct compositionality since it—like the standard account of Principle B—is
stated both non-locally and makes use of indices. But the standard account
will turn out to have many of the same problems as were discussed above with
respect to Principle B; co-indexation will not rule out the coreferential inter­
pretation in many cases and so a transderivational principle such as (48) will
be required. Thus regardless of one's feelings about direct compositionality,
one should be suspicious of this kind of account. My hunch is that Principle C
effects will derive from constraints on the packaging of information in dis­
course, as has been argued by, among others, Kuno (1987). If this is the case,
these have nothing at all to do with the compositional syntax and semantics
(unlike Principle B effects which—if the account above is correct—are located
in the semantic/syntactic composition). But this is left for another day.

References

CULY, C. 1985. 'The complexity of the vocabulary of Bambara', Linguistics and Philo-
sophy, 8: 345–51.
DAHL, Ö. 1974. 'How to open a sentence: abstraction in natural language', in Logical
Grammar Reports, 12, University of Gothenburg.
DEN DIKKEN, M., MEINUNGER, A., and WILDER, C. 2000. 'Pseudoclefts and ellipsis',
Studia Linguistica, 54: 41–89.
DOWTY, D. 1980. 'Comments on the paper by Bach and Partee', in J. Kreiman and
A. Ojeda (eds), Papers from the Parasession on Pronouns and Anaphora. Chicago:
Chicago Linguistic Society. 29–40.
— 1982. 'Grammatical relations and Montague grammar', in P. Jacobson and G. K.
Pullum (eds), The Nature of Syntactic Representation. Dordrecht: Reidel. 79–130.
(Reprinted by Springer, Classic Titles in Linguistics, 2006.)
— 1985. 'On recent analyses of the semantics of control', Linguistics and Philosophy,
8: 291–331.


Pardee, B. and Bach, E. 1981. ‘Quantification, pronouns, and VP anaphora’, in J.
Groenendijk, T. Janssen, and M. Stokhof (eds), Formal Methods in the Study of Lan-
guage: Proceedings of the Third Amsterdam Colloquium. Amsterdam: Mathematisch
Centrum. 445–81.
Pollard, C. 1984. Generalized Phrase Structure Grammars, Head Grammars, and
—— 2000. ‘Strategies of anaphora resolution’, in H. Bennis, M. Everaert, and E. Reu-
land (eds), Interface Strategies. Amsterdam: Royal Netherlands Academy of Arts and
Sciences. 295–324.
Language. Dordrecht: Reidel. 70–126.
Richardson, M. Marks, and A. Chukerman (eds), Papers from the Parasession on the
Interplay of Phonology, Morphology, and Syntax. Chicago: CLS. 198–221.
Schlenker, P. 2003. ‘Clausal equations (a note on the connectivity problem)’, Natural
—— 2005. ‘Non-redundancy: Towards a semantic reinterpretation of binding theory’,
Sharvit, Y. 1999. ‘Connectivity in specificational sentences’, Natural Language Seman-
Shieber, S. 1985. ‘Evidence against the context-freeness of natural language’, Linguistics
and Philosophy, 8: 333–43.
and Language, Revfülöp, Hungary.
Vijay-Shanks, K., Weir, D., and Joshi, A. 1986. ‘Adjoining, wrapping, and headed
strings’, in Proceedings of the 24th Meeting of the Association for Computational
Linguistics. New York: ACL.