Strict and Sloppy Reflexives in VP Ellipsis

Matthew Ong
Adrian Brasoveanu

This paper reports a series of three experimental studies (described in detail in Ong 2013) with three related goals/results. The first goal is to empirically evaluate two competing – syntax vs. discourse driven – accounts of strict vs. sloppy readings of reflexives in VP ellipsis, building on the experimental investigation in Kim and Runner (2009). The results strongly suggest that discourse-based accounts are empirically more adequate. The second goal is to argue that a heretofore ignored lexical factor, namely the meaning of the ellided verb, is in fact a strong predictor of strict vs. sloppy readings. We found that 'implicit-causality' verbs that are object-oriented are much more likely to have strict readings than subject-oriented implicit-causality verbs. Finally, we observe that the position of sentential negation is also an important biasing factor with respect to strict vs. sloppy readings, which we attribute to its 'reversal' function in discourse.

Keywords: reflexives, ellipsis, implicit causality, discourse relations, negation

1 Introduction

Consider the examples of VP ellipsis in (1a) and (1b) below. These sentences involve two clauses where the first one (the *source clause*) provides the antecedent for the elided VP in the second one (the *target clause*).

(1) a. John blamed himself, and Bill did too.
b. John blamed himself because Bill did too.

The VP ellipsis *did too* in the target clause is resolved to the VP *blamed himself* in the source clause, and the anaphor *himself* at the ellipsis site is ambiguous between a *strict* reading 'Bill blamed John' and a *sloppy* reading 'Bill blamed Bill, i.e. himself'.

The availability of strict vs. sloppy readings is affected by the choice of sentence coordinator/subordinator: it is more difficult to get a strict reading for (1a), while (1b) allows it more readily. This difference crucially involves VP ellipsis since when the target clause has a full overt VP (*Bill blamed himself*), Principle A of Binding Theory rules out strict readings quite strongly irrespective of what sentence coordinator/subordinator we use. Similar asymmetries between the availability of strict vs. sloppy readings can be found with a number of other subordinating conjunctions, for example, *even though*, *when*, and *before*.

We are grateful to Christina Kim who generously provided the original materials for her and Jeff Runner's experiments, as well as helpful email discussion of various issues. We are similarly indebted to Hannah Rhode for discussion and experimental data. Finally, we are grateful to Pranav Anand, Donka Farkas, Matt Wagers, and an anonymous *Empirical Issues in Syntax and Semantics 10* reviewer for extensive and very helpful discussions and/or comments that sharpened our interpretation of the experimental results and our theoretical hypotheses. The usual disclaimers apply.

Empirical Issues in Syntax and Semantics 10, ed. Christopher Piñón, 251–268
http://www.cssp.cnrs.fr/eiss10/

The present paper investigates this interpretation asymmetry between (1b) and (1a), hereafter called the Causality Effect. Specifically, we compare two competing explanations. One comes from Hestvik (1995) and is syntactic in nature, while the other is based on the theory of discourse relations in Kehler (2002). The paper also asks whether the asymmetry can be influenced by the semantics of the verb in addition to the syntactic and/or discourse relation between the two clauses. Since subordinating clauses (exemplified by *because*) often involve causality relations, it is worth asking whether *implicit causality* in the verb's meaning itself can trigger the same strict/sloppy bias. Finally, we investigate the perhaps unexpected role of sentential negation in biasing towards strict vs. sloppy readings.

We investigate these issues in a series of three binary-choice experiments that look at multiple aspects of the Causality Effect. These experiments are an extension of previous work by Kim and Runner (2009), which centered around the effect of discourse connectives on strict vs. sloppy readings of reflexives. In their work, the focus was on *parallelism* vs. *cause-effect* discourse relations in sentences such as *Mary voted for herself, and/so Jane did too.* The experiments reported in the present paper (described in detail in Ong 2013) expand on this idea in three ways: (i) we expand the range of discourse connectives while controlling for syntactic configuration; (ii) we specifically address the role of the verb's semantics in facilitating the Causality Effect; finally, (iii) we examine the role of negation in affecting the strict vs. sloppy bias.

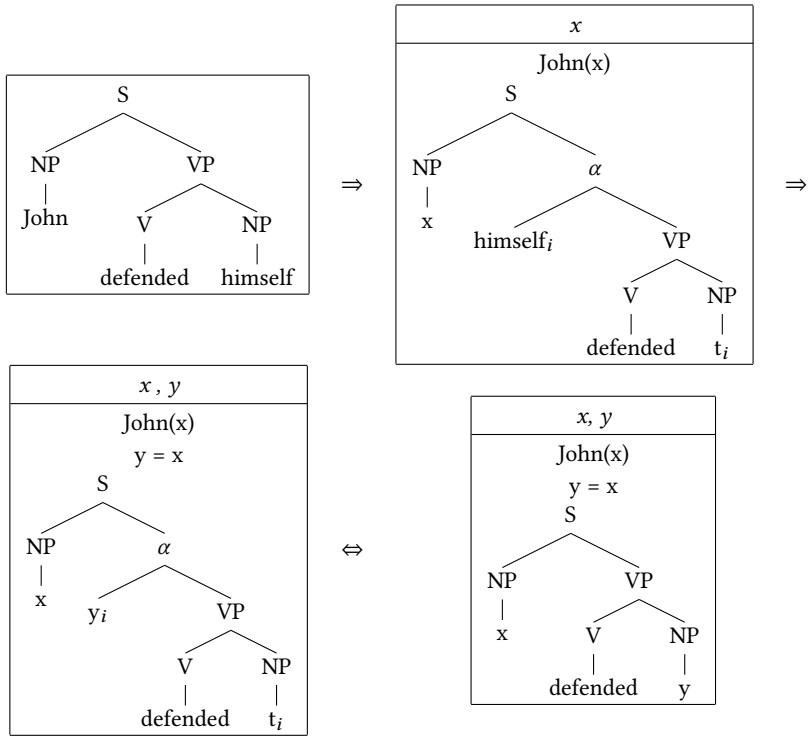
The main results are as follows. Experiments 1 and 3 strongly suggest that Kehler's (primarily) discourse-driven theory is a better model of the Causality Effect: the discourse relation between source and target clauses, and not their syntactic configuration, seems to be the major factor in determining how likely strict/sloppy readings are. Experiments 1 and 2 show that lexically-contributed causality plays a role in the availability of strict readings even when such causality is not specifically marked in the discourse relation. Finally, Experiment 3 also shows that sentential negation influences how likely strict vs. sloppy readings are; we conjecture that this is due to its 'reversal' function in discourse.

The paper is structured as follows. Sections 2 and 3 present the background and specific proposals for the syntactic and discourse-theory explanations of the Causality Effect, and Section 4 discusses the notion of *implicit causality*. Section 5 presents the three experimental studies. The first study was designed to test which theory, Hestvik's or Kehler's, makes better empirical predictions. The second study is a follow-up that focuses specifically on the role implicit causality plays in biasing toward strict or sloppy readings. Finally, the third study expands on the first by investigating additional connectives and their interaction with sentential negation. Finally, section 6 concludes.

2 A Syntactic Account: Hestvik (1995)

To explain why subordinate structures license strict readings but parallel/coordination structures do not, Hestvik proposes a copy-based account of VP ellipsis (cf. the deletion account in Sag 1976). The account is couched in a DRT framework (see Kamp 1981 and Kamp and Reyle 1993, and also Heim 1982 for a very closely related framework), and takes reflexives to receive their interpretation by undergoing LF movement out of the VP in order to establish the equational condition needed to resolve their anaphoric requirement (see Lebeaux 1983 and Chomsky 1986 among others for similar movement-based accounts). Movement essentially creates a λ-abstraction configuration that leads to a bound variable interpretation – see Figure 1.

Figure 1
The DRT account of reflexives in Hestvik (1995): '⇒' indicates that a DRS is derived from another by applying various syntactic and/or construction rules; ⇔ indicates semantic equivalence.



The derivation of strict vs. sloppy readings comes from competing orders between (i) the LF copying of the VP material to the ellipsis site and (ii) the raising of the reflexive out of the source VP. If raising happens before copying, both the trace in the source VP and the trace in the elided VP are governed by a single reflexive, giving the strict reading. If raising happens after copying, the trace at the ellipsis site is governed locally within its own clause, giving the sloppy reading. For example, the derivation of the sloppy reading for *John blamed himself because Bill did* proceeds as follows:

(2) John [$_{VP}$ blamed himself] because — Bill *e*
John [$_{VP}$ blamed himself] because — Bill [$_{VP}$ blamed himself]
John [$_{\alpha}$ himself$_i$ [$_{VP}$ blamed t$_i$]] because — Bill [$_{\alpha}$ himself$_j$ [$_{VP}$ blamed t$_j$]]

Recall that under this account, reflexives can be successfully interpreted only if they are able to move 'under' a suitable NP at LF. No movement, hence no binding, takes place before the VP is copied in (2) above. Since movement and variable binding happen only after copying, *Bill* ends up serving as the antecedent of the reflexive in the elided VP.

We derive the strict reading if we reverse the order of the two covert LF operations: we first raise the reflexive out of the VP, and copy the VP only after that.

(3) John [[$_{VP}$ blamed himself] because — Bill *e*]
John [$_{\alpha}$ himself$_i$ [[$_{VP}$ blamed t$_i$] because — Bill *e*]]
John [$_{\alpha}$ himself$_i$ [[$_{VP}$ blamed t$_i$] because — Bill [$_{VP}$ blamed t$_i$]]]

Crucially, the structure in the last line of (3) is licit because both traces t_i are bound by the reflexive *himself*$_i$, which is possible under the assumption that the *because* clause is adjoined below the subject of the source clause. When we try to follow the same derivation for the strict reading in the parallel configuration, the subject of the source clause does not c-command the target clause and the second trace t_i in in (4) below ends up being unbound and incurring an ECP violation:

(4) John [$_{\text{VP}}$ blamed himself] and Bill *e*
John [$_\alpha$ himself$_i$ [$_{\text{VP}}$ blamed t$_i$]] and Bill *e*
John [$_\alpha$ himself$_i$ [$_{\text{VP}}$ blamed t$_i$]] and Bill [$_{\text{VP}}$ blamed t$_i$]

One interesting issue Hestvik mentions in passing concerns the effect of negation on strict vs. sloppy readings, which will be the focus of our Experiment 3. He observes that when the source clause is negated, strict readings are available:

(5) John didn't blame himself, but Bill did. (strict reading possible)

This observation is unaccounted for under (the simple version of) the syntactic account: the presence of negation in the source clause does not affect the overall, coordination-based syntactic structure. So if syntactic structure was the primary determiner of strict vs. sloppy bias, sentences like (5) should behave like the sentence without negation in (4).

3 A (Primarily) Discourse-Relation Based Account: Kehler (2000, 2002)

Kehler (2000, 2002) proposes an alternative, discourse-relation based explanation for the Causality Effect. Under this account, there are three basic discourse relations, **Resemblance**, **Cause-Effect**, and **Occasion**, each with different 'subtypes', for example:

(6) a. Bill likes to play golf. Al likes surfing the net. (Resemblance: Parallelism)
b. John supports Clinton, but Mary opposes him. (Resemblance: Contrast)
c. Bill was about to be impeached. He called his lawyer. (Cause-Effect: Result)
d. Bill called his lawyer, (Cause-Effect: Explanation)
because he was about to be impeached.
e. Bill was about to be impeached, (Cause-Effect: Violated Expectation)
but he didn't call his lawyer.

Generally, a Resemblance relationship between S_1 and S_2 requires a one-to-one correspondence between the set of entities mentioned in S_1 and the set of entities in S_2, as well as some salient property (or more generally, relation) P that holds of both sets.

For Cause-Effect relationships, however, one need only have an implicational relationship between sentences at the propositional level. Here 'implicational' is defined in terms of plausibility and not in the logical sense of material or strict implication. Thus, if P is inferred from S_1 and Q from S_2, the Result relation is obtained if P plausibly 'implies' Q. Similarly, if Q plausibly 'implies' P, we have Explanation, and if P plausibly 'implies' $\neg Q$, we have Violated Expectation.

What is crucial in Kehler's theory is that for VP ellipsis, Resemblance relations require syntactic identity while Cause-Effect relations require identity only at the 'propositional level'. This distinction is meant to capture a wide range of observations about the acceptability of various perturbations of the prototypical examples of VP ellipsis, such as voice-mismatch (7),

and Condition C effects (8).

(7) In March, four fireworks manufacturers asked that the decision be reversed, and on Monday the ICC did. (from Dalrymple 1991, p. 35)

(8) Sue defended John$_i$ because he$_i$ couldn't. (based on examples from Kehler 2000, p. 550)

Kehler makes the same sort of observation about Condition A effects, where Cause-Effect relations license strict readings much better than Resemblance relations. For example:

(9) a. John$_i$ defended himself$_i$ even though Bill didn't. (Denial of Preventer)
b. John$_i$ defended himself$_i$ and so Bill did too. (Result)
c. John$_i$ defended himself$_i$ but Bill nevertheless didn't. (Violated Expectation)

4 Implicit Causality (IC)

In addition to the contribution of syntactic and discourse structure to the Causality Effect, the third important aspect considered in this paper is the lexical contribution of the verb. In particular, different verbs have different *implicit causality* (IC) biases (Garvey and Caramazza 1974, McKoon et al. 1993, Rohde 2008, Solstad and Bott 2013 and references therein). For example, the verbs *disappoint* and *scold* strongly bias pronoun resolution in distinct ways because of the cause-effect structures they are prototypically associated with. To see this, consider the minimal pairs below:

(10) John disappointed Bill because he (=John) stole the book.

(11) John scolded Bill because he (=Bill) stole the book.

Readers of (10) strongly prefer to resolve the pronoun *he* to the subject of *disappoint* rather than the object, whereas in (11) the opposite is true. The strength of the preference is so strong that Garvey and Caramazza (1974) think it is due to the fact that verbs like *disappoint* and *scold* imply as part of their root meaning an underlying causal event involving either the subject or object. For example, in *John disappointed Bill*, John must have done something to make Bill disappointed in him, while in *John scolded Bill*, Bill must have done something to make John scold him. Verbs in the class of *disappoint*, such as *amaze*, *infuriate*, and *frighten* have been dubbed IC1 verbs, and those in the class of *scold*, such as *thank*, *fear*, and *hate*, are known as IC2 verbs.

(12) List of IC verbs (from McKoon et al. 1993):
a. IC1: aggravate, amaze, amuse, annoy, apologize, bore, charm, cheat, confess, deceive, disappoint, exasperate, fascinate, frighten, humiliate
b. IC2: assist, blame, comfort, congratulate, correct, detest, envy, hate, jeer, notice, pacify, praise, reproach

What is significant about IC verbs is the possibility that they may trigger the Causality Effect in a way akin to discourse connectives like *because* or *even though*. That is, they induce a weakening of the requirement for structural parallelism that Kehler's model predicts for Resemblance relations. With such weakening, the elided VP can function like a deep anaphor (in the sense of Hankamer and Sag 1976), bypassing the 'structural identity' requirement associated with

Resemblance.

Since the choice of verb is (largely) independent of the type of discourse connective used, we predict that the Causality Effect induced by IC verbs should appear with both *and*-type and *because*-type connectives. Thus, even *and*-type sentences should exhibit a higher percentage of strict readings when IC verbs are present. This prediction is investigated in our Experiment 2, the results of which are reported in subsection 5.2.

5 The Three Experimental Studies

5.1 Experiment 1

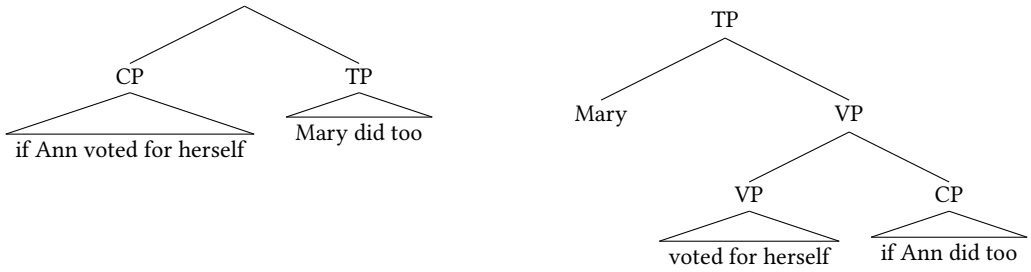
5.1.1 Reflexives and if*-clauses* As a first step towards deciding whether Hestvik's syntactic account or Kehler's discourse account better explains the Causality Effect, we designed an experiment that varied the relative c-command relation between *the subject of the source clause* and *the elided VP in the target clause* while preserving the discourse relation between the two clauses. The two configurations we used were conditional sentences in which the antecedent appeared either before or after the consequent:

(13) If Ann voted for herself, Mary did too. ('if-then' conditional)

(14) Mary voted for herself if Ann did too. ('then-if' conditional)

The reason for using conditional structures (a novel contribution to the experimental literature on this topic, as far as we can tell) is that their syntactic structure is fairly well understood, and the literature seems to be in agreement that the syntactic structures of 'if-then' and 'then-if' conditionals differ in exactly the respect we want (see e.g. Chierchia 1995 among others). In particular, 'if-then' conditionals have roughly the structure depicted in Figure 2 on the left, where the *if*-clause is adjoined higher than the main-clause subject, while the *if*-clause is adjoined at the VP level in 'then-if' conditionals, as shown on the right.

Figure 2
Syntactic structures for 'if-then' (left panel) and 'then-if' (right panel) conditionals.



A number of tests indicate that a sentence-final *if*-clause is adjoined below the subject of the matrix clause:

(15) *Condition C effects*:

a. *She$_i$ yells if Mary$_i$ is hungry. (*coreferential matrix-subject pronoun)

b. Bill visits her$_i$ if Mary$_i$ is sick. (✓coreferential matrix-object pronoun)

(16) *VP ellipsis*: I will leave if you do, and John will [~~leave if you do~~] too.

(17) *VP topicalization*: I told Peter to take the dog out if it rains, and [take the dog out if it rains] he will. (from Iatridou 1991, p. 9)

Hestvik's account predicts that 'if-then' conditionals should have only sloppy readings since the derivation of strict readings would require the subject of the source clause to c-command the elided VP. In contrast, both strict and sloppy readings are predicted to be possible for 'then-if' conditionals. Kehler's coherence account predicts that both strict and sloppy readings should be possible for either type of conditionals since the cause-effect relationship between the *if*-clause and the matrix clause is preserved regardless of linear order. These predictions are summarized in (18).

(18) Predicted readings:

	Structural account	Coherence account
If P, Q ('if-then' conditional)	sloppy (only)	strict & sloppy
Q if P ('then-if' conditional)	strict & sloppy	strict & sloppy

5.1.2 Method The experiment had a 2×3 factorial design, crossing 3 connectives (AND, IF, and SO) and the relative order of the source and target clause. An example item passed through all the conditions is provided in (19) below:

(19) Experiment 1 – example item:

	EARLY (generalizes 'if-then')	LATE (generalizes 'then-if')
AND	Ann voted for herself, and Mary did too.	Mary voted for herself, and Ann did too.
IF	If Ann voted for herself, Mary did too.	Mary voted for herself if Ann did too.
SO	Ann voted for herself, so Mary did too.	Mary voted for herself, so Ann did too.

The SO-conditions were included so that the results could be directly compared to the results reported in Kim and Runner (2009). Their Experiment 3 had a 2×2 factorial design crossing the Resemblance/Cause-Effect discourse relations and intra-/inter-sentential configurations, as exemplified in (20):

(20) Kim and Runner (2009), Experiment 3 – example item:

	RESEMBLANCE	CAUSE-EFFECT
INTRA-SENT.	Ann voted for herself and Mary did too.	Mary voted for herself so Ann did too.
INTER-SENT	Ann voted for herself. Mary did too.	Mary voted for herself. So Ann did too.

Although the main goal of Kim and Runner (2009) was to see if inter- vs. intra-sentential relations affected strict and sloppy readings for reflexives, we included the SO-conditions in our experiment to compare our results against their Resemblance vs. Cause-Effect manipulation. Many of the same verbs were used in both experiments.

In addition to the EARLY VS. LATE and CONNECTIVE TYPE manipulations, we paid particular attention to verb type. The verbs were chosen to be a mixture of implicit causality and non-causality verbs in order to see whether implicit causality (in either direction) had any effect

on strict vs. sloppy readings in any one of the conditions. In particular, 9 IC1, 18 IC2, and 21 NON-IC verbs were chosen.

The participants were given a binary choice task in which they were asked to (implicitly) choose between a strict and a sloppy reading in the context of a 'detective story'. The participants assumed the role of a police chief that was the boss of a 'concise detective', and were instructed to select the most likely interpretation of a report made by the detective about some on-going investigation. An example stimulus is provided below:

(21) The Detective reported to you: 'If Becky voted for herself in the election, Samantha did too.'
You understand this to mean that: If Becky voted for herself in the election, Samantha voted for
a. Becky
b. Samantha

The choice that the participants were required to make effectively disambiguated between the strict and sloppy interpretation of the detective's report.

This particular setup was chosen so that both conditional and AND/SO stimuli could be accommodated. That is, we could have followed Kim and Runner (2009) and simply ask the question: *Who did Samantha vote for? (A) Becky or (B) Samantha.* This would have been natural for AND/SO stimuli (*Becky voted for herself in the election, and/so Samantha did too.*), but this type of question would have been less natural for IF stimuli like the one exemplified in (21) above. This is because the question *Who did Samantha vote for?* ostensibly presupposes that Samantha actually voted, while the detective's report explicitly marks the *conditional*, uncertain status of this proposition.

31 UC Santa Cruz undergraduate students participated in the experiment for course (extra) credit. All participants were native speakers of English. The experiment was conducted using an installation of Alex Drummond's Ibex platform[1] locally hosted on the UCSC servers.

There were 48 experimental items and 60 fillers, 6 of which were control fillers used to assess whether participants were paying attention to the experimental task and did not select answers arbitrarily. Every participant saw each item exactly once; the items were rotated through the 6 conditions (Latin square design). The order of the 108 stimuli (48 experimental items + 60 fillers) was randomized for each participant, and the order of the two choices associated with each stimulus was randomized for every stimulus and every participant.

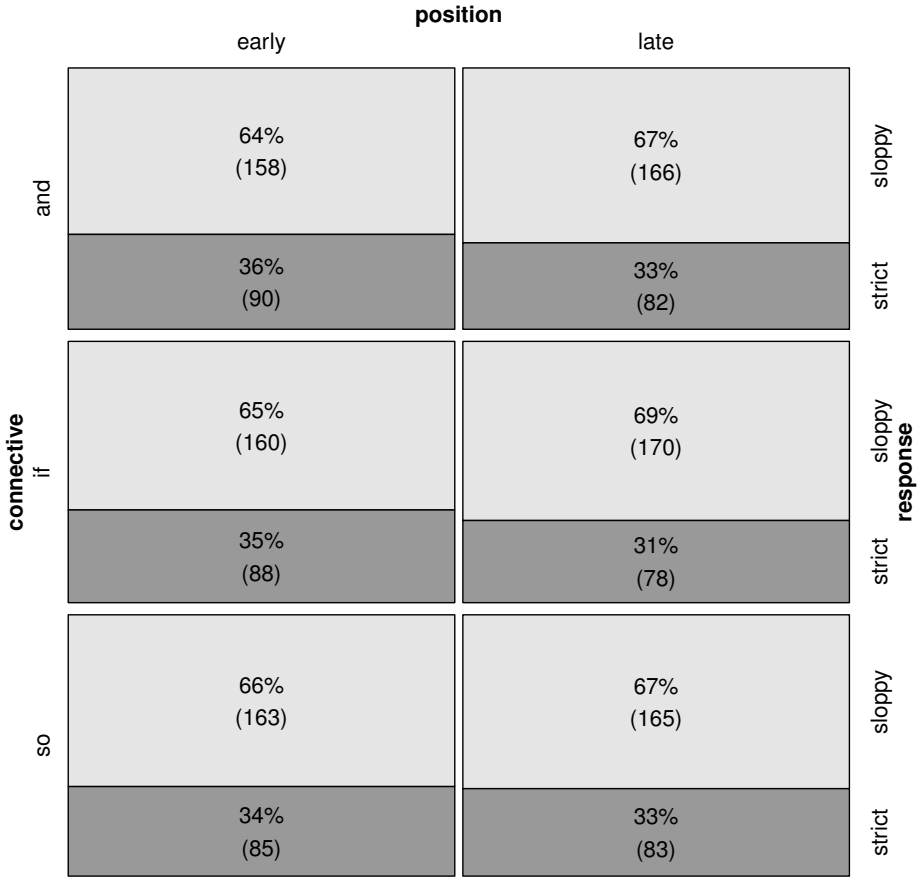
5.1.3 Results and discussion The percentages of strict/sloppy readings for the 6 conditions, followed by the raw counts in parentheses, are provided in Figure 3. Three generalizations can be extracted from these results. First, the percentage/probability of strict readings is roughly constant across all conditions. This is confirmed by the main-effects pnly and the interaction mixed-effect logistic regression models for this data (both models included crossed subject and item random intercepts and random slopes for connectives[2]): none of the effects were significant in either the main-effects or the interaction model, and the interaction model did not significantly reduce deviance compared to the main-effects model. This across-the-board null result

[1]See http://code.google.com/p/websprr/.

[2]This was the maximal random effect structure that converged; see Barr et al. (2013) for more discussion of (maximal) random effect structures for mixed-effects regression models.

is particularly significant for the IF-conditions: the order of the *if*-clause relative to the main clause does not appear to make any difference. This is compatible with the coherence account but not with the syntactic one – the latter predicts that there should be a significant difference between these two conditions.

Figure 3
Experiment 1: Percentages and raw counts of strict/sloppy for **connective**×**position**; the areas of the 6 boxes and of the strict/sloppy subregions inside each of them is proportional to the relative number of observations in that cell.



Second, contrary to what one might expect from isolated intuitive judgments, participants chose the strict reading fairly frequently for all connective types, even AND. This undermines both the account in Hestvik (1995) and the one in Kehler (2002) since both of them argue that under certain conditions, the reflexive in VPE should behave just as it would in the overt counterpart.[3] But Kehler's account, with its explicit acknowledgment of the multiple factors contributing to VPE licensing and interpretation, seems to be more easily generalizable to accommodate this result.

Third, our Experiment 1 and Experiment 3 in Kim and Runner (2009) suggest different conclusions about the impact of Cause-Effect relations on strict vs. sloppy readings in VPE. In Kim and Runner (2009), Cause-Effect showed a markedly higher tendency (>70%) toward strict readings, while Resemblance showed a probability of strict readings similar to ours.

[3]We are grateful to an anonymous reviewer for emphasizing this point.

A much less uniform picture of the facts emerges if we examine the data by verb, that is, by IC type. As Figure 4 shows, verb/IC type makes a clear contribution to the Causality Effect. This contribution is statistically significant: adding verb type as a third fixed effect to the mixed-effects models estimated above improves data fit ($\chi^2 = 5.25, df = 2, p = 0.07$), with a significant difference between IC1 and IC2 ($p = 0.02$) and a close-to-significant difference between IC1 and NON-IC ($p = 0.09$).

Figure 4
Experiment 1: Percentages and raw counts of strict/sloppy for **connective**×**verb type**; the areas of the 9 boxes and of the strict/sloppy subregions inside each of them is proportional to the relative number of observations in that cell.

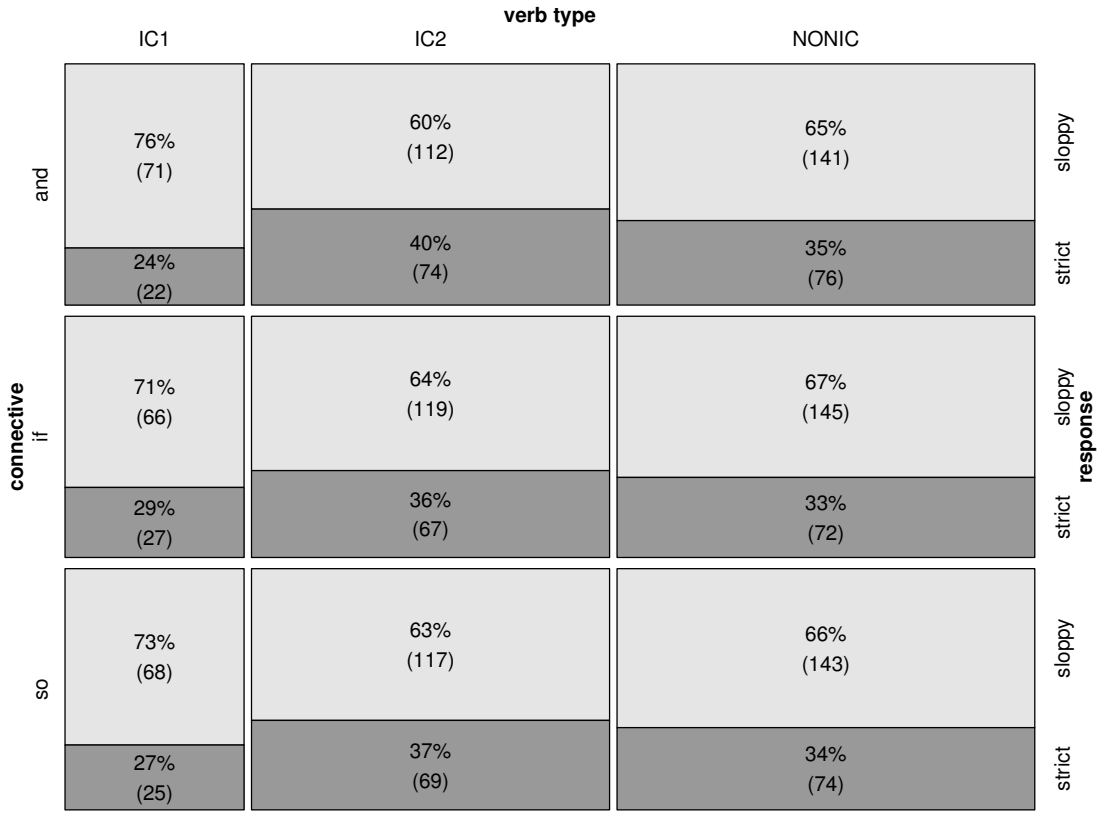


Figure 4 shows that the proportion of strict readings for IC1 is lower than for IC2, with NON-IC somewhere in between (but close to IC2). This holds for all connective types, suggesting that the phenomenon is at least partially independent of discourse relation.[4]

If as Kehler (2002) suggests, VP ellipsis is anaphora to properties (see also Hardt 1999 and Stone and Hardt 1999), the resolution of which is guided by a variety of factors (discourse structure, syntactic structure, etc.), then it is reasonable to expect that one of the factors biasing the resolution of this anaphoric requirement is the IC content contributed by the main verb. We already know that this IC content can strongly bias the resolution of regular pronouns (Rohde 2008 and references therein, among others), so it is plausible that it could have an impact on

[4]Since the effect of verb/IC type was very similar for the two clause orders (EARLY vs. LATE), we aggregated over them in Figure 4 for the sake of readability.

the resolution of higher-order anaphora.

One way to flesh this idea out would be to say that reflexives like *himself* can receive two distinct (but closely related) interpretations: (i) the default/preferred one is the 'de-transitivizer' interpretation – they are simply functions from transitives/binary relations to intransitives/unary relations, and (ii) their other, less salient interpretation is a pronominal one, with a suitably constrained resolution. See Büring (2005, 2011), Schlenker (2005), Jacobson (2007), Roelofsen (2008), and references therein, among many others, for more discussion of this and of the related preference for binding rather than coreference exhibited by pronouns.

The conjecture, then, is that IC1 vs. IC2 verbs interact with these two interpretations of reflexives in different ways. The subject-oriented IC1 verbs are compatible with the preferred de-transitivizer interpretation, and maybe even reinforce it. We therefore expect a conditional like *If John disgraced himself, Bill did too* to exhibit a strong preference for sloppy readings. If the overt reflexive *himself* in the antecedent is interpreted as a de-transitivizer – because it is the preferred interpretation of reflexives and because IC1 verbs highlight their subject and consequently, the remainder of the sentence is 'understood' as a predication about the subject – the covert reflexive in the elided VP will likely receive the same de-transitivizer interpretation, which will yield the sloppy reading.

In contrast, the object-oriented IC2 verbs highlight the object and its causal efficacy (see e.g. Kasoff and Lee 1993), and indicate that the sentence should be 'understood' as predicating a relation between the subject and the object. This would make the second, pronoun-like interpretation of the reflexive more salient since the de-transitivizer interpretation is explicitly not relational in nature. Now suppose the overt reflexive *himself* receives a pronominal interpretation in a conditional like *If John criticized himself, Bill did too.* Then this gives the subject *John* in the antecedent clause an extra salience boost and in addition, makes it more likely that the covert reflexive in the elided VP will receive the same pronominal interpretation. Since pronoun-like elements (whether overt or covert) prefer to retrieve the most salient entity, the likelihood of retrieving *John* in the consequent, that is, the likelihood of a strict reading, is higher.

5.2 Experiment 2

5.2.1 Motivation In order to directly investigate the effects of IC type (rather than indirectly by treating IC as a covariate, which is what we did in Experiment 1), we conducted a follow-up experiment that expanded the number of IC1 and IC2 verbs under investigation while eliminating NON-IC verbs. In this experiment, IC type was an explicit experimental manipulation. The synonym classes of both IC types were expanded, and since Experiment 1 already established that Kehler's discourse-based account of the Causality Effect is the more plausible one, the number of connective types was pared down to just AND and SO, and there was no manipulation of syntactic structure.

5.2.2 Methods and materials The experiment had a 2×2 design that crossed IC type (IC1 vs. IC2) and connective type (AND vs. SO). The IC verbs from Experiment 1 were reused along with a number of new verbs, which added up to 24 IC1 verbs and 24 IC2 verbs:

(22) List of verbs tested in Experiment 2:

a. IC1: amuse, disappoint, scare, humiliate, disgrace, encourage, motivate, reassure, fool, calm, inspire, embarrass, confuse, please, shock, startle, let down, flatter, amaze,

discourage, disgust, astonish, cheat, surprise

b. IC2: hate, pity, like, dislike, thank, help, condemn, congratulate, apologize, comfort, value, criticize, blame, berate, disparage, laugh at, correct, be disappointed with, be hard on, have confidence in, praise, defend, doubt, respect

There were 48 items and 60 fillers, and the procedure was identical to the one used in Experiment 1. There were 21 participants in this experiment, all of them UC Santa Cruz undergraduate students completing the experiment for course (extra) credit. An example IC1 item is provided in (23) (in the AND-condition), and an example IC2 item is provided in (24) (in the SO-condition).

(23) The Detective reported to you: 'Kevin amused himself by telling funny stories and Mike did too.'
You understand this to mean that: Kevin amused himself by telling funny stories and Mike amused
a. Kevin b. Mike

(24) The Detective reported to you: 'Cathy blamed herself for the disaster so Sally did too.'
You understand this to mean that: Cathy blamed herself for the disaster so Sally blamed
a. Cathy b. Sally

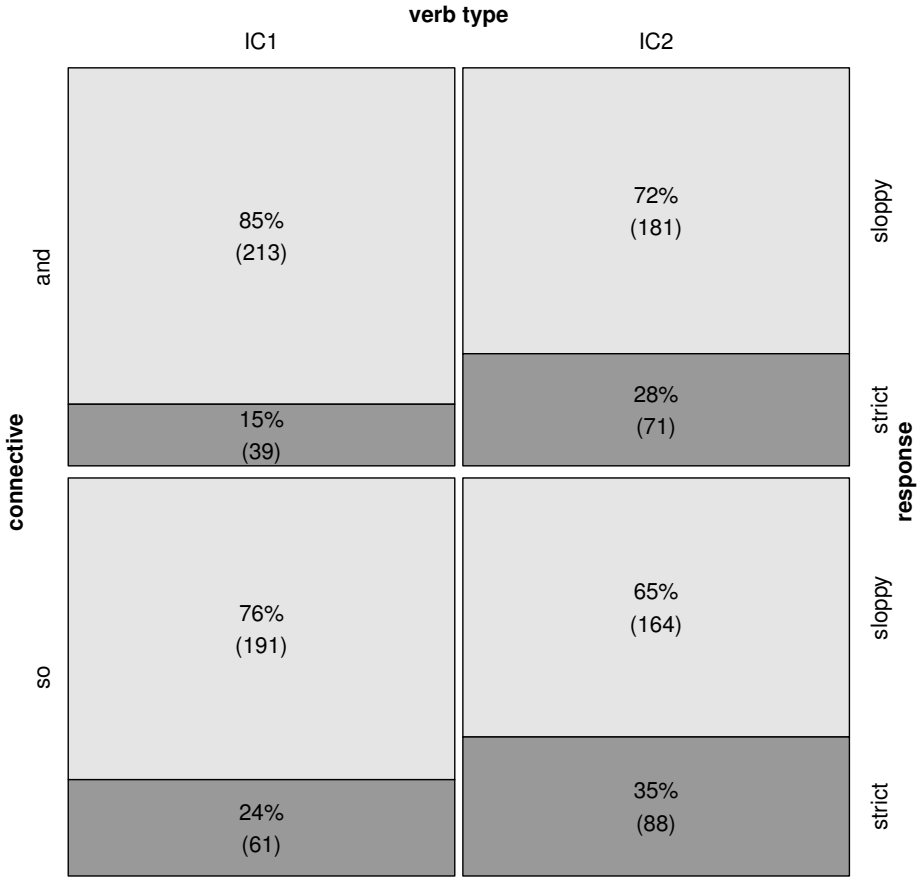
5.2.3 Results and discussion The results of Experiment 2 are summarized in Figure 5. The results confirm the observation that IC1 verbs have a depressing effect on the probability of strict readings, while IC2 verbs raise that probability.

This effect is highly significant. We compared two mixed-effects logistic regression models, one with main effects only for connective and verb type, and one with the interaction of connective and verb type in addition to the main effects. Both models had the maximal random effect structure possible for both subjects and items, that is, random intercepts and random slopes for the main effects of connective and verb type, and for their interaction.

The interaction model did not significantly improve fit relative to the main-effects only model. But in the main-effects model, both the effect of connective and the effect of verb type were highly significant (effects reported in logits; AND is the reference level for the connective fixed effect, and IC1 is the reference level for the verb-type fixed effect): $\beta_{\text{SO}} = 0.67, SE = 0.24, p = 0.005$ and $\beta_{\text{IC2}} = 0.82, SE = 0.27, p = 0.003$. Thus, we see that IC2 verbs have a strong positive effect on the probability of strict readings, independently of the enhancing effect of the connective SO – and stronger than this connective effect.

These results confirm and solidify our tentative results from Experiment 1. However, the same question from Experiment 1 persists regarding the relatively low proportion of strict readings for the SO conditions. We would expect higher proportions – along the lines of what Kim and Runner (2009) report – if SO really signaled a Cause-Effect relation. One possibility is that the particular experimental task we selected (because we wanted to investigate conditionals in Experiment 1) had a depressive influence on the probability of strict readings. One of the goals of our third study, to which we turn in the next section, was to investigate if this depressive effect can be observed with other Cause-Effect connectives in addition to SO.

Figure 5
Experiment 2: Percentages and raw counts of strict/sloppy for **connective**×**verb type**; just as before, the areas of the 4 boxes and of the strict/sloppy subregions inside each of them is proportional to the relative number of observations in that cell.



5.3 Experiment 3

5.3.1 Motivation The third, and final, study had two main goals. One of them, mentioned above, was to investigate if the particular experimental task we used had an overall depressive effect on the probability of strict readings for Cause-Effect connectives other than so. The second goal was to investigate the interaction between this broader range of discourse connectives and negation, following up on the observation in Hestvik (1995) that coordinating structures involving *but* and negation preferrably have a strict reading:

(25) John didn't blame himself, but Bill did.

Hestvik accounts for this phenomenon within his syntactic framework by arguing that *but* structures can sometimes involve subordination. However, the argument against this analysis is parallel to the argument against analyzing *and*-clauses as subordinated: neither can be fronted in the same way as other, clearer cases of subordinating conjunctions:

(26) a. Because Bill blamed himself, John blamed himself.
b. *but Bill blamed himself, John didn't blame himself.

c. *and Bill blamed himself, John didn't blame himself.

Given that discourse relations seem to be the main factor in influencing strict vs. sloppy readings, it is worth asking instead what impact the discourse structure associated with *but* + negation has on them. Is it primarily the negation that makes strict readings more likely? Or is it the contrast relation contributed by *but*? Or maybe both?

One way of (partially) distinguishing between these possibilities is to manipulate the position of negation: if we see an effect when the negation is in the first clause (and *but* follows it), rather than when the negation is in the second clause, we can more confidently say that negation has an important role (maybe in conjunction with *but*).

Similarly, if we see a systematic effect of the position of negation (first vs. second clause) across a variety of connective types in addition to *but*, we can more confidently attribute the effect to negation and its contribution to discourse structure, rather than attributing it to the contribution made by sentence connectives (or IC type, for that matter).

5.3.2 Method To test this, we used a 2×4 factorial design (plus 1 control condition that was identical to one of the conditions in Experiments 1 and 2) that crossed 4 discourse connectives – AND, BUT, (AND) THEREFORE, and (BUT) NEVERTHELESS – and the presence of negation in either the first or the second clause. The extra control condition was AND with no negation. An example item is provided below:

(27) Experiment 3 – example item (NTL=nevertheless, TF=therefore):
[Control – AND & NO NEGATION: John blamed himself and Bill did too.]

	EARLY NEGATION	LATE NEGATION
AND	John didn't blame himself and Bill did.	John blamed himself and Bill didn't.
BUT	John didn't blame himself but Bill did.	John blamed himself but Bill didn't.
NTL	John didn't blame himself but nevertheless Bill did.	John blamed himself but nevertheless Bill didn't.
TF	John didn't blame himself and therefore Bill did.	John blamed himself and therefore Bill didn't.

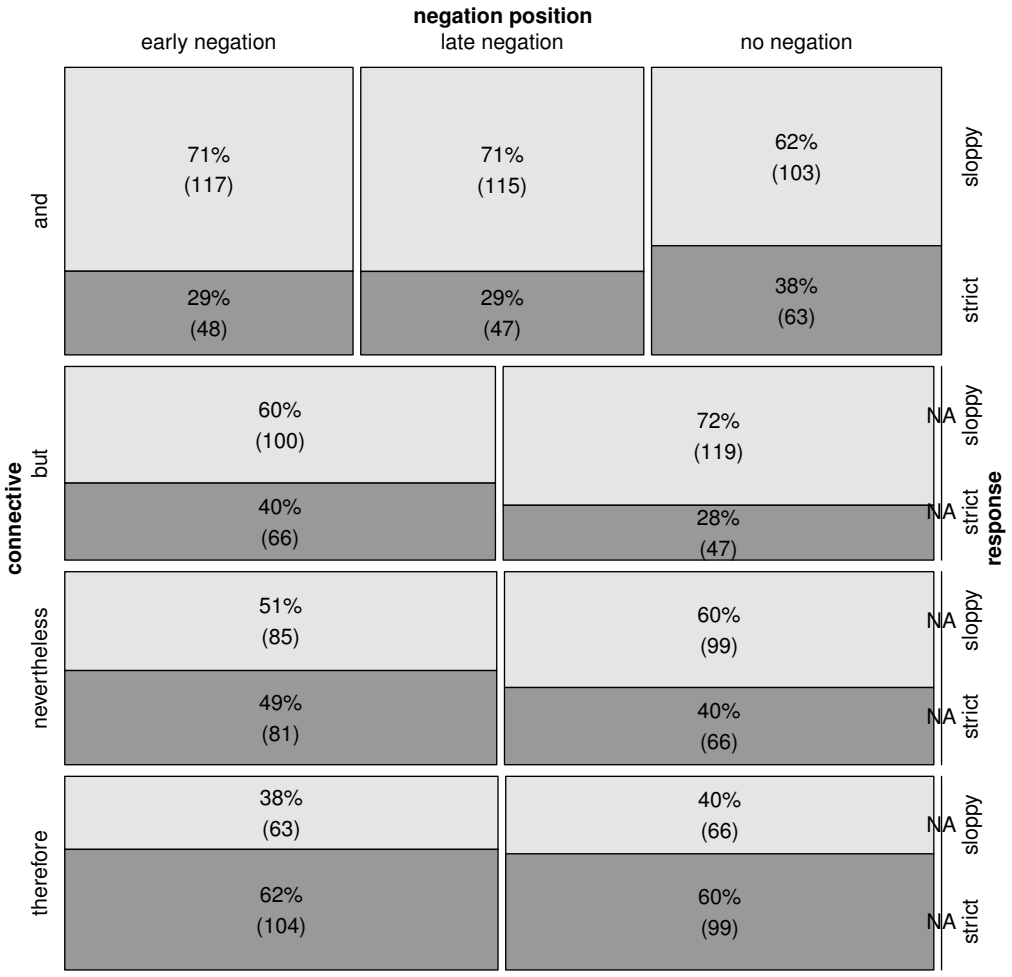
Experiment 3 used the same items and fillers as Experiment 1, and the experimental procedure also remained the same. 31 UC Santa Cruz undergraduate students participated in this experiment for course (extra) credit.

5.3.3 Results and discussion Figure 6 provides the descriptive summary of the Experiment 3 data. We see that the percentage of strict readings (38%) we obtained for the control condition (AND & NO NEGATION is comparable with the percentages we obtained for the same condition in Experiment 1 (36% for EARLY and 33% for LATE, with a non-significant difference between EARLY and LATE). This indicates that the overall nature of the task was very similar across the two experiments, so we can draw conclusions about the experimental task in general based on the results of our Experiment 3.

The results show that the experimental task does not have an across-the-board depressive effect on the probability of strict readings: the percentages of strict readings for THEREFORE and NEVERTHELESS are high, and very close to the ones observed in Kim and Runner (2009) for

so. Thus, while our specific task might depress the probability of strict readings, the depressive effect is not as strong as to swamp all effects of connective type. It is therefore possible that the lack of difference between AND and SO in Experiment 1 and the fairly small difference between them in Experiment 2 are due to the particle SO, which might not be an unambiguously Cause-Effect expressing particle. Instead, SO might express a more general, semantically bleached discourse relation that subsumes Cause-Effect, for example, some type of weak notion of plausible entailment or 'relatedness'.

Figure 6
Experiment 3: Percentages and raw counts of strict/sloppy for **connective×negation position**; just as before, the areas of the boxes and of the strict/sloppy subregions inside each of them is proportional to the relative number of observations in that cell; the six NAs in the figure mark unavailable cells, that is, the combinations of conditions that were not tested: NO NEGATION & BUT, NO NEGATION & NEVERTHELESS, and NO NEGATION & THEREFORE.



To analyze the data, we compared two mixed-effects logistic regression models, one with main effects only for connective type and negation position, and one with interaction terms between connective type and negation position in addition to the main effects. Both models had the maximal random effect structure that converged for both subjects and items (and within

those non-nested maximal models, the smallest deviance), namely: random intercepts and random slopes for connectives.

The interaction model did not significantly improve fit relative to the main-effects only model. But in the main-effects model, the effect of connective for NEVERTHELESS and THEREFORE (but not for BUT), and the effect of negation were highly significant (effects reported in logits): $\beta_{\text{NEVERTHELESS}} = 0.81, SE = 0.25, p = 0.001$, $\beta_{\text{THEREFORE}} = 1.63, SE = 0.28, p = 3 \times 10^{-9}$, and $\beta_{\text{LATE-NEGATION}} = -0.32, SE = 0.12, p = 0.01$.

We see that both THEREFORE and NEVERTHELESS have a significantly higher probability of strict readings than AND (the reference level for the connective fixed effect) or BUT. This provides further support for the coherence account since it is not at all clear that clauses headed by *therefore* or *nevertheless* are syntactically subordinated.

Importantly, we also see that LATE NEGATION has a significantly lower probability of strict readings than EARLY NEGATION (the latter being the reference level for the negation-position fixed effect). Although this happened to some extent across all connectives, it manifested itself most strongly for BUT and NEVERTHELESS.

One way to explain the enhancing effect of early negation on the probability of strict readings is to follow Krifka (2013) (and references therein; see also Horn 1989) and take negative sentences to be verum-focused, or more precisely, to always contribute (or retrieve) a propositional discourse referent for their positive counterparts. That is, a negative sentence is not simply an assertion that happens to be negative, but crucially involves rejecting its positive counterpart. Negation has a reversal discourse function.

Since early negation makes its positive counterpart salient, the subject of the first clause receives a 'double boost' in salience, since it is part of both the asserted negative sentence and its positive alternative. The extra salience boost increases the likelihood of strict readings by a reasoning similar to the one we used to explain the Causality Effect observed in Experiments 1 and 2. As an anonymous reviewer points out, this seems to be closely related to the fact that examples like *John blamed himself but nobody else did* and *John blamed himself and everybody else did too* seem to strongly facilitate strict reflexives in a way that is hard for Hestik's or Kehler's accounts to capture; see Fiengo and May (1994), p. 105, fn. 10, and Kennedy (2003), p. 32 et seqq. for related discussion.

The contrastive function of BUT and NEVERTHELESS might work off of and reinforce the reversal effect associated with early negation. When these two connectives are used, the second clause (i.e. the clause immediately following BUT and NEVERTHELESS) is more likely to have a strict reading because it is expected to contrast with the first, negative clause, and therefore elaborate on the positive alternative evoked by that clause. See, for example, Vicente (2010), for more discussion of the 'corrective' use of *but* that involves a denial of the proposition expressed by the first conjunct (e.g. *John didn't go to the park, but (rather) he went to the library*).

6 Conclusion

The overall theme of this paper has been that discourse structure plays a significant biasing / disambiguating role with respect to strict vs. sloppy readings of reflexives in VP ellipsis. And this discourse structure is determined by a variety of sources, including the specific discourse connective that is used, the early vs. late position of sentential negation, and the semantics of the verb itself.

While we initially framed the theoretical contribution of the paper in terms of distinguish-

ing between two competing theories that were grounded in syntax vs. discourse structure, the results indicate that the empirical landscape is more fine-grained and complex than an antithesis of two simple sources of bias. The preponderance of the evidence seems to favor a (primarily) discourse-structure account, but not all of the data can be easily explained by examining discourse connectives and their meanings. For one thing, it is still unclear what the status and contribution of SO is. But more importantly, much of the explanatory burden was ultimately shifted to the meaning contributions made by other items: the implicit causality bias contributed by different verb types, the various analyses of reflexives proposed in the previous literature, and finally, the reversal (*verum*-focus related) contribution made by sentential negation. These effects and their interactions were merely outlined here, but they deserve a much more in-depth empirical and theoretical investigation.

References

Barr, Dale J., Roger Levy, Christoph Scheepers and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. In *Journal of Memory and Language* 68, 255-278.

Bhatt, Rajesh and Roumyana Pancheva. 2006. Conditionals. In *The Blackwell companion to syntax*, Martin Everaert and Henk van Riemsdijk (eds.), Blackwell, 638-687.

Büring, Daniel. 2005. *Binding theory*. Cambridge University Press.

Büring, Daniel. 2011. Pronouns. In *Semantics: An International Handbook of Natural Language Meaning, vol. 2*, Klaus von Heusinger, Claudia Maienborn and Paul Portner (eds.), De Gruyter, 971-996.

Chierchia, Gennaro. 1995. *Dynamics of meaning: anaphora, presupposition, and the theory of grammar*. University of Chicago Press.

Chomsky, Noam. 1986. *Knowledge of language: its nature, origin, and use*. Praeger.

Dahl, Östen. 1973. On so-called sloppy identity. In *Gothenburg Papers in Theoretical Linguistics* 11, University of Gothenburg (Sweden).

Dalrymple, Mary. 1991. Against reconstruction in ellipsis. *System Sciences Laboratory Report* SSL-91-114, Xerox PARC.

Fiengo, Robert and Robert May. 1994. *Indices and identity*. MIT Press.

Garvey, Catherine and Alfonso Caramazza. 1974. Implicit causality in verbs. *Linguistic Inquiry* 5, 459-464.

Hamkamer, Jorge and Ivan Sag. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7, 391-426.

Hardt, Daniel. 1999. Dynamic interpretation of verb phrase ellipsis. *Linguistics and Philosophy* 22, 187-221.

Heim, Irene. 1982. *The semantics of definite and indefinite noun phrases*. PhD dissertation, UMass Amherst. [Published in 1988 by Garland Publishing]

Hestvik, Arild. 1995. Reflexives and ellipsis. *Natural Language Semantics* 3, 211-237.

Horn, Larry R. 1989. *A natural history of negation*. CSLI.

Iatridou, Sabine. 1991. *Topics in conditionals*. PhD Thesis, MIT.

Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar*. MIT Press.

Jacobson, Pauline. 2007. Direct compositionality and variable free semantics: the case of 'Principle B' effects. In *Direct Compositionality*, Chris Barker and Pauline Jacobson (eds.), Oxford University Press, 191-236.

Kamp, Hans. 1981. A theory of truth and semantic representation. In *Formal Methods in the Study of Language*, Jeroen Groenendijk, Theo Janssen and Martin Stokhof (eds.), Mathematical Centre Tracts 135, 277-322.

Kamp, Hans and Uwe Reyle. 1993. *From discourse to logic*. Kluwer.

Kasoff, Joseph and Ju Young Lee. 1993. Implicit causality as implicit salience. *Journal of Personality and Social Psychology* 58, 765-778.

Kehler, Andrew. 2000. Coherence and resolution of ellipsis. *Linguistics and Philosophy* 23, 533-575.

Kehler, Andrew. 2002. *Coherence, reference, and the theory of grammar*. CSLI.

Kennedy, Christopher. 2003. Ellipsis and syntactic representation (logical and phonological). In *The In-*

terfaces: Deriving and Interpreting Omitted Structure, Kerstin Schwabe and Susanne Winkler (eds.), Amsterdam: Benjamins.

Kim, Christina S. and Jeffrey T. Runner. 2009. Strict identity, coherence, and parallelism in VP ellipsis. In *Proceedings of Semantics and Linguistic Theory 19*, Satoshi Ito and Ed Cormany (eds.), CLC, 275-287.

Krifka, Manfred. 2013. Response particles as propositional anaphors. In *Proceedings of Semantics and Linguistic Theory 23*, Todd Snider (ed.), CLC, 1-18.

Lebeaux, David. 1983. A distributional difference between reciprocals and reflexives. *Linguistic Inquiry* 14, 723-730.

Ong, Matthew. 2013. *Strict and sloppy reflexives in VP ellipsis*. MA thesis, UC Santa Cruz.

McKoon, Gail, Steven Greene and Roger Ratcliff. 1993. Discourse models, pronoun resolution, and the implicit causality of verbs. *Journal of Experimental Psychology* 19, 1040-1052.

Roelofsen, Floris. 2008. *Anaphora resolved*. PhD dissertation, ILLC, University of Amsterdam.

Rohde, Hannah. 2008. *Coherence-driven effects in sentence and discourse processing*. PhD dissertation, UC San Diego.

Sag, Ivan. 1976. *Deletion and logical form*. PhD dissertation, MIT. [Published in 1980 by Garland Publishing]

Schlenker, Phillipe. 2005. Non-redundancy: towards a semantic reinterpretation of binding theory. In *Natural Language Semantics* 13, 1-92.

Solstad, Torgrim and Oliver Bott. 2013. Towards a formal theory of explanatory biases in discourse. In *Proceedings of the 19th Amsterdam Colloquium*, Maria Aloni, Michael Franke and Floris Roelofsen (eds.), ILLC, 203-210.

Stone, Matthew and Daniel Hardt. 1999. Dynamic discourse referents for tense and modals. *Second International Workshop on Computational Semantics (IWCS II)*, Harry Bunt, Reinhard Muskens, and Gerrit Rentier (eds.), Tilburg University.

Vicente, Luis. 2010. On the syntax of adversative coordination. *Natural Language and Linguistic Theory* 28, 381-415.

(Ong)
UC Santa Cruz
matthewcong@gmail.com

(Brasoveanu)
UC Santa Cruz
abrsvn@{gmail.com,ucsc.edu},
http://people.ucsc.edu/~abrsvn/