NPI Licensing, Strawson Entailment, and Context Dependency

KAI VON FINTEL
MIT

Abstract

The Fauconnier–Ladusaw analysis of negative polarity licensing (that NPIs are licensed in the scope of downward entailing operators) continues to be the benchmark theory of negative polarity. In this paper, I consider some of the moves that are needed to maintain its basic intuition in some recalcitrant arenas: negative polarity licensing by *only*, adversatives, superlatives, and conditionals. We will see that one has to (i) use a notion of entailment that I call *Strawson Entailment*, which deals with presuppositions in a particular way, and (ii) prohibit (even natural) context change during an inference. The purpose of this paper is to investigate how to justify these constraints and to see in detail how the semantics of the problematic constructions has to work in order for these moves to successfully rescue the Fauconnier-Ladusaw analysis.

I will first show the two assumptions at work in the analysis of NPI licensing by *only* and adversatives (building on proposals by Kadmon & Landman). I then turn to NPI licensing in the antecedent of conditionals. The standard Stalnaker–Lewis semantics for conditionals—*if p, q* is true iff *q* is true in the *closest p*-world(s)—might make one suspect that once one has an explanation for NPI licensing by superlatives, that would immediately deliver an explanation for NPI licensing in conditionals. But it turns out that the particular analysis that seems appropriate for NPI licensing by superlatives cannot plausibly carry over to conditionals. Instead, one does better by appealing to an alternative analysis of conditionals, one that I have elsewhere argued for on independent grounds.

I BASICS

Negative polarity items (NPIs) are expressions that can prototypically occur in the semantic scope of negation but not in 'positive' environments. Two examples are *any* and *ever*:

(1) a. I **don't** think we have <u>any</u> potatoes.
         #I think we have <u>any</u> potatoes.
     b. I **don't** think there will <u>ever</u> be another Aristotle.
         #I think there will <u>ever</u> be another Aristotle.

Since NPIs occur not just in the scope of negation but also in a variety of other 'affective' environments (the term is due to Klima 1964), one needs a theory of NPI licensing. Building on an idea from Fauconnier (1975, 1979),

Ladusaw (1979, 1980a, b) argued for a semantic theory of NPI Licensing whose core claim is that NPIs are licensed in 'downward entailing' (DE) positions. This analysis has proven to be the benchmark theory of NPIs and as such has of course been attacked (especially by Linebarger 1980, 1987, 1991; Israel 1996) and refined (Hoeksema 1983, 1986a; Krifka 1990, 1991b, 1994, 1995; Kadmon & Landman 1993; van der Wouden 1994; Jackson 1995; Zwarts 1995, 1997; Giannikidou 1997). In this paper, I will be concerned with some recalcitrant arenas of NPI licensing. My strategy will be to see how far one can push the Fauconnier–Ladusaw approach.

The basic intuition of the approach is that NPI licensing expressions share with negation the property of reversing the direction of entailment in their argument slot. First, observe that negation clearly has this property.

(2) It rained hard. ⇒ It rained.
     It is not the case that it rained. ⇒ **It is not the case that** it rained hard.

Assume, uncontroversially, that *it rained hard* entails *it rained*. When negation takes these sentences as its argument, the entailment is reversed: *It is not the case that it rained* entails *it is not the case that it rained hard.*

For expressions that do not take sentences as their argument, one needs to generalize the notion of entailment before one can test whether the expression reverses the direction of entailment. Quantificational determiners (like *some, every, no,* etc.) can be taken as making claims about the relationship between two sets of individuals (supplied by the common noun argument of the determiner and the rest of the sentence). *Some* says that the two sets have at least one member in common, *no* says that they have no members in common, and *every* says that the first set is a subset of the second. Say that for two sets A and B, A 'entails' B iff A is a subset of B. (The intuition is simply that for any given individual x, *x is in A* entails *x is in B*). A determiner reverses entailment (is 'downward entailing', from sets down to subsets) in a given one of its argument slots iff we can replace a set in that slot with any of its subsets while preserving truth. By this criterion, *some* is not DE in either of its argument positions (in fact, it is 'upward entailing', UE), *no* is DE in both of its slots, while *every* is DE in its first slot and UE in its second slot.

(3) sparrow ⇒ bird
     Some sparrow is in the tree. ⇒ **Some** bird is in the tree.
     No bird is in the tree. ⇒ **No** sparrow is in the tree.
     Every bird is in the tree. ⇒ **Every** sparrow is in the tree.
(4) is chirping ⇒ is making noise
     Some sparrow is chirping. ⇒ **Some** sparrow is making noise.

No sparrow is making noise. ⇒ No sparrow is chirping.
Every sparrow is chirping. ⇒ Every sparrow is making noise.

Quite spectacularly, we find that NPI licensing exactly mirrors these entailment properties. *Some* does not license NPIs in either of its arguments, *no* licenses them in both of its arguments, and *every* licenses NPIs in its first argument but not in its second argument:

(5) **Some** (student who has *ever been to Rome) (bought *any postcards there)
No (student who has ever been to Rome) (bought any postcards there)
Every (student who has ever been to Rome) (bought *any postcards there)

By the way, we see in (5) that the power of NPI licensing reaches down into the relative clause modifying the noun inside the first argument of a determiner.

Technicalities

Permit me to wax technical for a moment. We define a cross-categorial notion of entailment:

(6) *Cross-Categorial Entailment* (⇒)
   a. For p, q of type t: p ⇒ q iff p = False or q = True.
   b. For f, g of type $\langle \sigma, \tau \rangle$: f ⇒ g iff for all x of type σ: f(x) ⇒ g(x).

The base case is given by the truth-function of material implication. One truth-value 'entails' another iff it is not the case that the first is True and the second is False. Generating higher-level notions of entailment is done by saying that one function 'entails' another of the same type iff for any argument of the appropriate type the result of combining the first function with the argument 'entails' the result of combining the second function with the same argument.

The most immediate application is to the notion of entailment between propositions, functions of type $\langle s, t \rangle$ from worlds to truth-values. One proposition p will entail another q iff for any world w, p(w) entails—by the base clause—q(w), that is, iff for all worlds either p is false or q is true. In set-talk, this will be the case iff the set of worlds for which p is true is a subset of the set of worlds for which q is true.

Another simple but important application is to 'entailment' between functions of type $\langle e, t \rangle$, that is between functions from individuals to truth-values. One such function 'entails' another iff for any individual, the truth-value obtained by applying the first function to the individual entails the truth-value obtained by applying the second function to the same individual. Since such functions can be seen as characteristic functions of

sets of individuals (those individuals for which the function yields the truth-value True), we can move to set-talk: one set 'entails' another iff the first is a subset of the second.

Next, we define DE:

(7) *Downward Entailingness* (DE)
   A function f of type $\langle \sigma, \tau \rangle$ is downward entailing (DE) iff for all x, y of type σ such that $x \Rightarrow y$: $f(y) \Rightarrow f(x)$.

A function is DE iff for any two arguments such that the first 'entails' the second, the result of applying the function to the second argument 'entails' the result of applying the function to the first argument.

The Fauconnier–Ladusaw proposal

(8) *NPI Licensing*
   An NPI is only grammatical if it is in the scope of an α such that $[[\alpha]]$ is DE.

Some immediate qualifications should be made clear. (i) The Fauconnier–Ladusaw condition in (8) is just a necessary condition. There may well be further conditions, in particular there may be conditions having to do with the kind of speech act intended with an utterance containing NPIs (cf. Linebarger 1980, 1987, 1991). (ii) Linebarger has shown further that it is not enough for the NPI to occur *somewhere* in the scope of a DE-expression. Rather, it must be in the 'immediate' scope. There must not be an intervening non-DE-operator. How exactly to circumscribe the relevant class of blocking interveners is an open issue (see Kadmon & Landman 1993 for discussion). (iii) The way I set this system up, the relevant DE-operator in *No boy bought any potatoes* is not the determiner *no* (which, however, would be the licenser for an NPI inside the common noun argument) but the whole NP *no boy*. This does not prevent us from specifying a derivative notion according to which the ultimate licenser of the NPI would be *no*.

Just to be clear: to check whether an expression f, of type $\langle \sigma, \tau \rangle$, is DE (and thus may be an NPI licensor), we need to check the validity of the following inference for any x, y of type σ:

(9) $x \Rightarrow y$
   $\therefore f(y) \Rightarrow f(x)$

To facilitate judgments, one should of course pick argument expressions for which the entailment in the premise does intuitively hold. Then, this premise may be left tacit, and we just have to elicit judgments of the kind: 'Does *every bird is in the tree* entail *every sparrow is in the tree*?'

The Fauconnier–Ladusaw analysis is (in the words of Linebarger 1987: 361) 'impressively algorithmic' and is worth defending against challenges. Trouble for it can come from two directions, one less worrisome than the other. Since the DE-condition is only intended as a necessary condition, there is plenty of work explaining why sometimes a DE-operator licenses NPIs and sometimes it does not. This project could pull the rug out from underneath the analysis, if it turns out that these additional factors can by themselves explain NPI licensing. And that is clearly Linebarger's perspective, one that I will have nothing to say about

More directly problematic for the Fauconnier–Ladusaw account are cases where NPIs like *any* and *ever* are clearly licensed but where it is also clear that simple DE inferences are not valid. Here, the supposed necessary condition for NPI licensing does not seem satisfied. Appeals to additional factors cannot be of any help here. I will be concerned with four such cases: licensing by *only*, adversative attitude predicates, superlatives, and antecedents of conditionals. What I will suggest is that there are a number of related package deals. To save the basic account we have to do some work on clarifying the relevant notion of entailment. This tailored notion of entailment then needs to be combined with particular semantic treatments of the four mentioned constructions. The most immediate benefit of this investigation is that the semantics of these constructions has to be more deeply explored. Even if the project is eventually judged to be doomed, one might hope that we have found out more about how these constructions work.

2 ONLY

2.1 *The Problem*

The first problematic item we will discuss is *only*.

(10) **Only John** <u>ever</u> ate <u>any</u> kale for breakfast.

(11) **Only John** <u>ate vegetables for breakfast.</u>
    ↛ **Only John** <u>ate kale for breakfast.</u>

The generalized quantifier *only John* clearly licenses negative polarity items in its immediate scope.[1] At the same time, *only John* is apparently not downward entailing. It is easy to imagine scenarios where the premise of

---

[1] We will not explore the interesting issue of NPI licensing in the *first* argument of *only*. As discussed in von Fintel (1997), while such cases do exist, they may best be analyzed as NPI licensing by a hidden generic operator under *only*.

(11) is true but the conclusion is not obviously true: imagine John ate spinach for breakfast and no one else had any vegetables for breakfast. Given the (true) piece of information that only John ate vegetables for breakfast, and no more information, we are certainly not going to conclude that only John ate kale for breakfast. The juxtaposition of the licensing fact illustrated in (10) and the failure of downward entailment in (11) is exactly the kind of problematic data for the Fauconnier–Ladusaw account that I wish to explore in this paper.

A first attempt

Ladusaw suggested the following move: since we are really supposed to be testing for *entailment*, i.e. a truth-preserving inference, we need to abstract away from *conventional implicatures*. He argued for the move first in the case of so-called implicative verbs (Ladusaw 1979: 160f.). Consider:

(12) John **failed** to buy any shirt.
    a. John **failed** to buy a shirt.
    b. ⇒? John **failed** to buy a red shirt.

Ladusaw assumes that *x fails to* Q strictly speaking *entails* only that x doesn't have the property Q. The rest of its meaning is a conventional implicature that there was an attempt or expectation that x would have the property Q. 'Since entailment depends only upon truth-conditional meaning, it will be true that (a) entails (b), even though that intuition is confused by the fact that (b) implicates something that is not implicated or entailed by (a). (a) implicates that John tried or was expected to buy a shirt, but (b) implicates that he tried to buy a red shirt. The implicature is irrelevant to the question of whether (a) entails (b).'

To establish the DE-ness of *fail*, we need to be convinced that in a situation where it is true that John failed to buy a shirt it is also strictly speaking true that he failed to buy a red shirt. It may be misleading and odd to actually assert the conclusion but that does not necessarily mean that it isn't *true*. One can perhaps agree with Ladusaw that this move is adequate in the case of *fail*. But does it work for *only*?

Ladusaw (1979: 165) appeals to Horn's famous 'asymmetric', presuppositional analysis of *only* (Horn 1969). The idea is that in (13), (a) asserts what (b) expresses and presupposes what (c) expresses:

(13) a. Only John loves Mary.
    b. No one who is not John loves Mary.
    c. John loves Mary.

The special character of the positive component of the meaning of *only* is revealed by a number of facts that look like presupposition projection (invariance/inheritance under negation, interrogation, and other embedding environments). Horn (1969) assumed that the positive component was a semantic presupposition, failure of a situation to verify this component would result in the *only*-statement being neither true nor false. In later work, various versions have been attempted according to which the status of the positive component is something other than a semantic presupposition. For example, Horn (1979) treats the positive component as a conventional implicature, a non-truth-conditional constraint on the appropriate assertion of an *only*-sentence. This is the kind of analysis that Ladusaw must assume for his move to work.

To establish the DE-ness of *only*, we need to be convinced that in a situation where it is true that only John ate vegetables for breakfast it is also strictly speaking true that only he ate kale for breakfast. It may be misleading and odd to actually assert the conclusion but that does not necessarily mean that it isn't *true*.

The problem with this account is that it is highly controversial that *only*-sentences can in fact be true in situations where their positive component is not satisfied. If one believed that the positive component is a presupposition in any semantically relevant sense, one would not assent to such an assessment. And if one thought that the positive component is in fact an entailment (albeit perhaps 'backgrounded' in some way), one would also decline to accept Ladusaw's pleading. And there are plenty of people in these camps (to appreciate the heat of debate one just has to look at Atlas 1993, 1996).

Notice also that quite apart from our theoretical allegiances, it just isn't good methodology to base a semantic theory on judgments about the truth of a sentence in a situation where it would be misleading and inappropriate to assert the sentence.

2.2 *The remedy*

Fortunately, we can improve on Ladusaw's proposal. We can define a notion of entailment that will only check whether an inference is truth-preserving *under the assumption* that all the conventional implicatures and presuppositions of premises and conclusion are satisfied. Assuming for now and for concreteness that all such presuppositions are to be treated as definedness conditions imposed by partial functions denoted by presupposition triggers, we can define a modified notion of downward entailment as follows.

(14) *Strawson Downward Entailingness*
    A function f of type $\langle \sigma, \tau \rangle$ is Strawson-DE
    iff for all x, y of type σ such that $x \Rightarrow y$ and $f(x)$ *is defined:* $f(y) \Rightarrow f(x)$.

(In a short while, I will discuss the name I gave to this notion of entailment). This new notion checks downward entailment with the additional premise that the conclusion has a defined semantic value.

Let's see how this would pan out in the case of *only*. Assume this semantics for *only*:[2]

(15) $[\![only]\!]$ (x) (P) is defined only if P(x) = True.
    If defined, $[\![only]\!]$ (x) (P) = True iff $\neg \exists y \neq x$: P(y) = True.

Now to figure out whether *only John* is downward entailing in the new sense, we need to check the validity of the following inference schema:

(16) $P \Rightarrow Q$
    $[\![only\ John]\!]$ (P) is defined
    $[\![only\ John]\!]$ (Q) = True

    ∴ $[\![only\ John]\!]$ (P) = True.

Since we assume that *only John is P* will be defined iff *John is P* is true, we can spell out the schema as follows:

(17) $P \Rightarrow Q$
    John is P
    Only John is Q

    ∴ Only John is P

Here is an example of what we need to check:

(18) Kale is a vegetable.
    John ate kale for breakfast.
    Only John ate vegetables for breakfast.

    ∴ Only John ate kale for breakfast.

---

[2] Actually, we should probably adopt Horn's amended proposal (Horn 1996, 1997), according to which the presupposition triggered by *only* is not that P(x) = True but that there is a y such that P(y) = True. In the simplest case, this presupposition together with the assertion of *only* (that nothing other than x is P) will derive the claim that P(x) = True. The two proposals come apart in more complex cases. Horn argues that the new proposal is preferable and I am convinced. Nevertheless, I stick to the more commonly known formulation since this will avoid unnecessary side-investigations.

Let me just demonstrate that adopting Horn's new analysis would not threaten our project here. *Only* will still be Strawson-DE: *Only John ate kale for breakfast* according to the new analysis presupposes that someone ate kale for breakfast. Now, with the additional premise that someone ate kale for breakfast, *only John ate vegetables for breakfast* will entail that *Only John ate kale for breakfast*.

Since by adding the premise that the conclusion's presuppositions are satisfied we ensure that we do not need to worry about judging truth-values in contexts where there are unsatisfied presuppositions, the task has become much cleaner. Of course, the inference in (18) is intuitively impeccable. If NPI licensing is sensitive to the notion of Strawson entailment, we have an explanation for the fact that noun phrases like *only John* license NPIs in their scope.

Strawson

The notion of DE entailment in (14) is linked to the following notion of validity:

(19) *Strawson-Validity*
An inference $p_1, \ldots, p_n \therefore q$ is *Strawson-valid* iff the inference $p_1, \ldots, p_n, S \therefore q$ is (classically) valid, where S is a premise stating that the presuppositions of all the statements involved are satisfied.

Here is why I named these notions after Strawson. Strawson (1952) in a famous passage discusses the possibility of making the traditional inference from *Every S is P* to *Some S is P* valid within a modern logical framework. He essentially proposes that natural language quantifiers carry an existence presupposition with respect to their domain. He understood this presupposition to be a semantic presupposition: if the presupposition is not satisfied, the sentence will be neither true nor false. In his system, *Every S is P* can only be true (or false even) if there are S. Now, if *Every S is P* is true it must therefore be true that there are S and that all of them are P. Hence, it will follow (as desired) from *Every S is P* that *Some S is P*.

But this move to a system with existence presuppositions threatens the validity of other equally traditional patterns. For example, conversion from *Every S is P* to *Every non-P is non-S* does not validly follow (the premise may well be true while there are no non-Ps at all, but such a circumstance would make the conclusion suffer from presupposition failure). Similarly, the inference from *No S is* P to *No P is S* might founder on presuppositional rocks (in a case where there are no Ps at all, the premise will be true while the conclusion would again suffer from presupposition failure). Strawson's remedy was to suggest that what validity in traditional logic meant is this: 'We are to imagine that every logical rule of the system, when expressed in terms of truth and falsity, is preceded by the phrase "Assuming that the statements concerned are either true or false, then . . ."' (Strawson 1952: 176). As far as I know, this notion of validity has not been investigated much. But it does turn up again and again in the literature that is concerned

with NPI licensing and downward entailment. I will provide citations as we go along.

Initial justification

Strawson Entailment seems like a very peculiar notion that should have no place in a serious logical semantics.[3] Surely, it is just a plain mistake to think that *Only John ate vegetables for breakfast* entails *(entails!) Only John ate kale for breakfast*. It is hard to imagine I could emulate the incredulity of the reaction by Atlas to the asymmetric analysis of *only*, cf. Atlas (1996: 280), so let me quote a representative passage:

Suppose Socrates and a cohort of young aristocratic Athenians are conversing in the Athenian agora. As Plato puts a question to Socrates, a small mangy animal (a dog, perhaps) pushes his head through the crowd of legs and catches the alert Socrates' eye. No one else sees the animal. Then *Only Socrates saw an animal* is true. On the downwards monotonic view, it is **entailed—entailed!!**—that *Only Socrates saw a Tibetan snow leopard*. That is, in every possible model (world) in which *Only Socrates saw an animal* is true, *Only Socrates saw a Tibetan snow leopard* is true in that model. Surely this condition cannot be satisfied. The logical consequence relation cannot squeeze more information out of a conclusion than has been put into the premises. But that is what downwards monotonicity does in this sentence. So much the worse for downwards monotonicity in *only a* sentences.

While there are, of course, valiant attempts to justify a semantics for *only* where the inference that so offends Atlas' sensibilities is perfectly truth-preserving, one cannot deny the force of the objection.

Note, however, that with the move to Strawson Entailment we are not saying that *Only John ate vegetables for breakfast entails Only John ate kale for breakfast*. The claim is that the first sentence *Strawson entails* the second. That means that the first *entails* the second *under the assumption* that the presuppositions of the second are satisfied. The inference from the first to the second is only legitimate as an *enthymematic* or *elliptical* inference, one that relies on an additional tacit premise, namely that the presupposition of the conclusion is satisfied. Strawson Entailment does not pretend to 'squeeze more information out of a conclusion than has been put into the premises', it just introduces a particular kind of elliptical inference. Given merely the truth of the overt premise, we will not move to the conclusion (there is no entailment). But once we additionally know that the presupposition of the conclusion is satisfied, we can infer the truth of the conclusion. And that seems unobjectionable.

(i) Consider a situation where you know that John ate kale for breakfast. Someone tells you that only John ate vegetables for breakfast. It is entirely

---

[3] In fact, such a notion is not mentioned in Corcoran's extensive survey of uses of the term 'implies' (Corcoran 1973).

within your logical rights, to say: 'That entails then that only John ate kale for breakfast.' (ii) Consider someone who tells you 'I just found out that only John ate vegetables for breakfast. That entails then that only John ate kale for breakfast.' You would not, I think, accuse this person of a logical blunder. You would sensibly assume that the person is taking for granted that John ate kale for breakfast, and that she is using this knowledge as a tacit premise for her inference.

Strawson Entailment is, I claim, a particularly useful way of describing a natural way of introducing tacit premises into natural reasoning. The thesis explored in this paper is that NPI licensing is sensitive to this notion.

Note that the move to Strawson-DE still shares with the original Fauconnier–Ladusaw analysis the 'algorithmic' character. Its adequacy depends on formulating the semantics of *only* in such a way that the positive component of the meaning of *only* can be detected as special and converted into an additional assumption in the DE inference. Of course, one might point out that the test inference in (18) is also considered valid by anyone defending a completely symmetric semantics for *only*. Couldn't a symmetricalist appeal to the validity of this inference just as well to explain NPI licensing by *only*? No. The symmetricalist has no principled way to skim off the positive component of the meaning of *only* and convert it into a premise in a DE inference. The Strawson-DE move depends on a modular asymmetric semantics for *only*.

Another example of Strawson-DE licensing of NPIs

Sabine Iatridou (pc) sketches another case in which Strawson-DE appears to license NPIs:

(20) **It's been five years since I saw a bird of prey in this area.**
   $\not\Rightarrow$ **It's been five years since I saw an eagle in this area.**

(21) **It's been five years since I saw <u>any</u> bird of prey in this area.**

(22) It's been five years since I saw a bird of prey in this area.
   Five years ago I saw an eagle in this area.
   $\therefore$ It's been five years since I saw an eagle in this area.

This construction is not downward entailing as the problematic inference in (20) shows. Nevertheless, (21) shows that NPIs are licensed by this construction. We observe that *it's been five years since p* asserts that *p* hasn't been true since five years ago and presupposes that *p* was indeed true five years ago. The Strawson-DE experiment in (22) works fine.[4]

---

[4] Iatridou explores this construction and related ones in English and Greek in as yet unpublished work.

2.3 *Atlas' alternative*

Horn (1996) revisits the semantics of *only* and argues that the NPI licensing facts show that *only* must be downward monotonic in some sense. He doesn't discuss very much at all the issue of how to precisely state the kind of downward monotonicity displayed by *only*. At one point (p. 18) he does say that *only* licenses downward inferences 'if we ignore existential import'. It is not clear what ignoring existential import would amount to. But he goes on: 'Thus if I eat meat only when I'm depressed, then it follows that in particular I eat pork (if I eat it at all) only when I'm depressed.' It seems then that Horn is close to realizing that *only* is downward monotonic with respect to a tacit additional premise. Horn also cites (in his fn. 26, p. 33) a similar move suggested by Hoeksema in work on NPIs licensed by superlatives, a topic to which we will turn later.

Atlas (1996) rejects Horn's asymmetric semantics for reasons that do not have directly to do with NPI licensing. He adopts a by and large symmetric (conjunctionalist) analysis of *only*. *Only John ate kale for breakfast* asserts both that John ate kale for breakfast and that no one other than John did so. The two components of the meaning are of the same nature. Hence it will not be possible to filter out one of them (the positive claim) by a tailored notion of entailment as we have proposed.

As an alternative to the modified Fauconnier–Ladusaw analysis of NPI licensing by *only*, Atlas offers the following observation: *only John* is 'pseudo-anti-additive.' Zwarts had introduced the notion of anti-additivity as a property of functions that is stronger than DE, in a theory that attempts to explain the licensing requirements of different kinds of NPIs.

(23) A function f of type $\langle \sigma, \tau \rangle$ is *anti-additive* iff for all x, y of type σ: f(x) & f(y) = f(x $\cup$ y), where $\cup$ is a cross-categorial disjunction.

It can be shown that all anti-additive functions are DE, but not the reverse.

Atlas observes that the equivalence in (23) holds in one direction for *only John*:

(24) a. Only John drinks and only John smokes
    $\Rightarrow$ Only John drinks or smokes.
  b. Only John drinks or smokes
    $\nRightarrow$ Only John drinks and only John smokes.

To see why, suppose that the premise in (b) is true because only John drinks and nobody at all smokes. Then, the conclusion will not be true because it is not true that only John smokes. (At least, one may reasonably think that *only John smokes* is not true in such a situation. If one believes in an analysis that treats the positive component as irrelevant to the truth of the *only*-claim,

one might claim that *only John* is fully anti-additive. But we are now working on the other side of the dogmatic schism.)

Atlas then suggests that the weak property of 'pseudo-anti-additivity' is what licenses NPIs in the scope of *only John.*

(25) A function f of type $\langle \sigma, \tau \rangle$ is *pseudo-anti-additive* iff for all x, y of type σ: f(x) & f(y) $\Rightarrow$ f(x $\cup$ y).

He writes that pseudo-anti-additivity may make *only John* 'semantically "negative" enough to license some Negative Polarity Items—though there is no theory to explain why it does' (p. 283f). In a footnote (fn. 23, repeated verbatim also as number (3) of the open questions in Appendix 1), he adds:

Of course there is the possibility that the semantical explanation for *only Proper Name* co-occurring with 'weak' NPIs is quite other than the semantical pseudo-anti-additivity of *only Proper Name.* The hypothesis that downwards monotonicity was necessary for licensing NPIs was a noble attempt, but at this moment there is no theory at all for the behavior of *only Proper Name.* The observation that the expression is closed under finite unions is just an observation that one DeMorgan relation is satisfied. If that explains why it licenses 'weak' NPIs, no one has a theory that shows why that is an explanation. The young man or woman who comes up with a good theory wins a US$500 Atlas Prize.

I have to disappoint Atlas.[5] Pseudo-anti-additivity *cannot* be what explains NPI licensing under *only John.* The property of pseudo-anti-additivity is one that *only John* shares with a host of other quantifiers some of which license NPIs while others don't.

(26) Other 'Pseudo-Anti-Additive' Noun Phrases
    Some student smokes and some student drinks
    $\Rightarrow$ Some student smokes or drinks.
    No student smokes and no student drinks
    $\Rightarrow$ No student smokes or drinks.
    Every student smokes and some student drinks
    $\Rightarrow$ Every student smokes or drinks.
    At least three students smoke and at least three students drink
    $\Rightarrow$ At least three students smoke or drink.

Of these, *no student* licenses NPIs in its scope, while *some student, every student,* and *at least three students* do not license NPIs in their scope.[6]

---

[5] This is not just a disappointment for Atlas' theory, but also one for my wallet. I could really have used that $500 Atlas Prize. :

[6] Note that the fact that *no* licenses NPIs and is pseudo-anti-additive is not a triumph for the proposal under discussion. Presumably the fact that *no* is indeed anti-additive and not just pseudo-anti-additive is what explains its licensing behavior. What's under debate is whether mere pseudo-anti-additivity has anything to do with NPI licensing.

Quantifiers that do not obey pseudo-anti-additivity include:

(27) *Some Noun Phrases that are not 'Pseudo-Anti-Additive'*
At most three students smoke and at most three students drink
$\not\Rightarrow$ At most three students smoke or drink.
Only three students smoke and only three students drink
$\not\Rightarrow$ Only three students smoke or drink.
All but two students smoke and all but two students drink
$\not\Rightarrow$ All but two students smoke or drink.
Exactly one student smokes and exactly one student drinks
$\not\Rightarrow$ Exactly one student smokes or drinks.

Most of these (perhaps all) in fact license NPIs in their scope. It is thus clear that pseudo-anti-additivity is neither a necessary condition nor a sufficient condition for NPI licensing. It is *useless* for the analysis of NPI licensing. The conjunctionalist will need to cast around for another way of explaining the NPI licensing behavior of *only*. In the absence of a good analysis of NPI licensing in the scope of *only*, it would be advisable to explore further the consequences of adopting Strawson Entailment as the operative notion in a semantic theory of NPI licensing.

Prospectus

This analysis can now be seen from a number of perspectives. If one likes the basic idea of Fauconnier–Ladusaw and accepts an asymmetric semantics of *only*, then one has a reason to explore why Strawson Entailment should be part of the theory of grammar. If one likes the basic idea of Fauconnier–Ladusaw and accepts Strawson Entailment as not too implausible, one has an argument for an asymmetric semantics of *only* vis-à-vis some competitors. If one finds Strawson Entailment utterly implausible, one has nascent arguments both against the basic idea of Fauconnier–Ladusaw and an asymmetric semantics of *only*. I will explore the move made here of adopting Strawson Entailment and what one would have to say about the modular semantics of some other NPI licensing environments. The case of *only* was meant as a first illustration of the move.

3 ADVERSATIVES

So-called adversative attitude predicates like *be amazed, be surprised, be sorry, regret* license negative polarity items in their complement clauses:

(28) a. Sandy is **amazed/surprised** that Robin <u>ever</u> ate kale.
    b. Sandy is **sorry/regrets** that Robin bought <u>any</u> car.

If Ladusaw was right, then the complement clauses of these predicates should be a position for downward inferences. Supposing that p entails q, *Sandy is amazed that q* should entail *Sandy is amazed that p*. But once again this is clearly not the case:[7]

(29) Robin ate kale $\Rightarrow$ Robin ate a green vegetable.
    Sandy is **amazed/surprised** that <u>Robin ate a green vegetable</u>.
    $\not\Rightarrow$ Sandy is **amazed/surprised** that <u>Robin ate kale</u>.
(30) Robin bought a Honda Civic $\Rightarrow$ Robin bought a car.
    Sandy is **sorry/regrets** that <u>Robin bought a car</u>.
    $\not\Rightarrow$ Sandy is **sorry/regrets** that <u>Robin bought a Honda Civic</u>.

Part of the problem will be the same as before: there are interfering presuppositions and a move to Strawson Entailment will provide a remedy. Adversative predicates are factive: they presuppose the truth of their complement (more precisely, they presuppose that the subject of the attitude believes that the complement is true; in the absence of special contexts, they routinely suggest that the complement is in fact true). So, downward inferences cannot be relied on to be truth-preserving, since the conclusion may suffer from presupposition failure. From *Sandy regrets that Robin bought a car* we cannot safely infer that *Sandy regrets that Robin bought a Honda Civic*, since for all we know Robin did not buy a Honda Civic (or at least Sandy may be unaware of Robin buying a Honda Civic).

We can see, then, that the move to Strawson Entailment would be helpful here as well. If Strawson Entailment is all we need for NPI licensing, what we have to check is whether the entailment in (30) holds under the additional premise that Robin did in fact buy a Honda Civic (and that Sandy is aware of that). The usefulness of this move has been noticed quite often. Here are some relevant citations:

Ladusaw (1980a) on dealing with factive presuppositions
    The factivity of such predicates as *regret* obscures their true DE nature. For the purposes of determining whether an environment is DE, we should look only at situations in which the presuppositions of the sentences in question are satisfied.

---

[7] There is one worry about talking about entailment relations in attitude contexts that I will entirely (although perhaps not legitimately) ignore: since attitude holders are not as a rule logically consistent, their belief systems will often not behave as they logically should. To avoid this problem, assume counterfactually that attitude holders are perfectly consistent. If NPI licensing by attitude predicates is due to downward entailment, it would seem that the grammar (if not its users) makes such counterfactual assumptions.

Linebarger (1987) reporting a personal communication from Ladusaw
> It will only be appropriate to evaluate these proposed entailments with respect to the worlds in which all the presuppositions and/or conventional implicatures of the *consequent* are satisfied.

Asher (1987) on DE inferences with adversatives
> Weakened Downward Entailment (WDE). An attitude predicate α is WDE iff the following entailment holds:
> a α that ϕ
> $[\![\phi]\!] \Rightarrow [\![\varphi]\!]$
> a believes that φ
> _____
> $\therefore$ a α that φ.

Katz (1991) on entailments with factive predicates
> S entails S' iff both S and S' have truth value and when S is true in a context, S' is as well.[8]

Let us assume then that the problems about the conclusion adding new information can be circumvented by moving to Strawson Entailment as the criterion for NPI licensing.

Adversatives are more complicated though. Even when we presuppose that Robin bought a Honda Civic (and that Sandy is aware of that), it does not seem obvious at all that from *Sandy regrets that Robin bought a car* we can safely infer that *Sandy regrets that Robin bought a Honda Civic*. After all, the following seems perfectly coherent:

(31) Sandy regrets that Robin bought a car, but Sandy does not regret that Robin bought a Honda Civic.

The story that goes with it may be that Sandy really didn't want Robin to have a car but among the cars Robin could have gotten, Sandy by far prefers the Honda Civic.

Now, if one wants to maintain that *regret* is DE, one has to explain the prima facie coherence of (31). At the same time, one has to ensure that the story does not in turn make predicates like *glad* DE in the same sense. The latter is a suspicion voiced by Linebarger: as soon as we apply all the tricks in the book to make *regret* and its cousins DE, the same fate befalls *glad* and *its* cousins.

3.1 *Kadmon & Landman's move*

Kadmon & Landman (1993), who develop a variant of the Fauconnier-Ladusaw analysis, respond to the problem raised by (31) as follows: What

_____

[8] Of course, this needs to be fixed up a little but the intent of defining something like Strawson Entailment is clear.

makes (31) coherent is a change of 'perspective' between the two conjuncts. On a *consistent* perspective, the DE inference will be valid. Here's what they have to say:

To be surprised that A is always relative to a certain perspective on A, a perspective that determines what it is about A that is surprising and in virtue of what it is surprising. The perspective is a contextually determined parameter in the interpretation of *surprised*, very much in the same way that a 'modal base' (Kratzer 1981) is a contextually determined parameter in the interpretation of modals. It can be at least partly specified by explicit linguistic text, as in (32) and (33), or be left entirely implicit.

(32) I was surprised that he stole the watch, in as far as that was a daring thing to do.

(33) Given my high opinion on his moral character, I was surprised that he stole the watch.

The perspective enters into the semantics of *surprised* and affects the truth conditions of sentences containing it. That is, for example, why we (normally) don't take B in (34) to express a contradiction.

(34) A: Were you surprised that he stole the watch?

B: I was and I wasn't.

B is saying that relative to one perspective she was surprised and relative to another perspective she wasn't.

Thus, while *surprised* . . . may superficially look like a two-place relation, it can be semantically regarded as a THREE-PLACE relation between a subject (experiencer), a contextual perspective, and a proposition. Given that, to see whether *surprised* is DE, we have to check whether the following pattern holds. (A, B propositions, x subject, p, q perspectives.)

If A $\Rightarrow$ B then Surprised(x, p, B) $\Rightarrow$ Surprised(x, p, A)

That is, we have to check whether *surprise* is DE 'on a constant perspective'. The apparent DE failure case of 'surprised by the purchase of a car but not by the choice of car' is not relevant, since it involves a different pattern, viz., Surprised(x, p, B) $\not\Rightarrow$ Surprised(x, q, A).

There is nothing prima facie objectionable about such an analysis. It is quite standard to assume that validity of inferences is to be checked against the background of a constant context. Violations of the constancy of context are branded as *fallacies of equivocation*. There are two reasons why I would like to explore this approach into much further detail than Kadmon & Landman do: (i) I want to know what the fine-grained semantics of these factive attitude predicates looks like and what this contextual parameter 'perspective' does. (ii) I am skeptical that 'constant context' is a notion that can be uncritically assumed: after all, we have plenty of reason to suspect that context continually changes. This is the essence of the myriad of *dynamic* approaches to interpretation. What are we to make of a notion of DE entailment that relies on a constant context? I suppose the best perspective on this would be to make a distinction between two kinds of discourse: evolving deliberative discourse and static statements of logical

inferences. Someone who puts forward an inference from any number of premises (including tacitly assumed ones) to a target conclusion arguably guarantees implicitly that the inference does not rely on any context changes midway through the argument.

But such large issues should be put on the backburner while we tackle the nitty-gritty. What is this contextual parameter 'perspective' that Kadmon & Landman appeal to and how does it enter into the semantics of the attitudes? Kadmon & Landman are not very specific in their discussion. Perhaps it would be best to expand on their reference to Kratzer's analysis of the context-dependency of modals.[9] In a Kratzer-style semantics, we would make attitude predicates sensitive to two contextually supplied parameters of interpretation. One parameter will be a set of worlds, another will be an ordering among these worlds. For attitudes like *want, wish, glad, regret, sorry* the ordering will be one of 'preference'. For attitudes like *expect, amazed, surprised* the ordering will be one of 'expectation/likelihood'.

Kadmon & Landman do not say whether we should think of their 'perspective' parameter as corresponding to something like the set of worlds among which an ordering is established or as corresponding to whatever supplies the ordering itself. Looking at the examples that they discuss, (32) and (33), one would have to say that it is the ordering that is being grounded: what makes his stealing the watch more surprising than the possibility that he didn't steal the watch are the facts that stealing is a daring thing to do or that I have a high opinion of his moral character.

But when one looks at the crucial Honda Civic scenario, it seems that what changes is not the ordering. Instead, there appears to be a narrowing of the set of worlds among which the ordering is specified. What makes (31) coherent appears to be that when one takes into consideration worlds in which Robin does not buy a car, Sandy prefers those over worlds in which Robin does buy a car. But when we narrow the view to worlds that differ only in which car Robin buys, Sandy prefers those in which Robin buys a Honda Civic. This game can of course be continued at infinitum:

(35) Sandy is sorry Robin bought a car. But Sandy is not sorry that Robin bought a Honda Civic. However, Sandy is sorry that Robin bought a Honda Civic EX. Then again, Sandy is not sorry that Robin bought a Honda Civic EX with a sunroof.

---

[9] An explicit attempt to apply Kratzer's notions to the semantics of *regret* and to thereby salvage DE can be found in Katz (1991). Instead of reporting directly what he has to say, I will unfold the analysis in a slightly different way.

So, perhaps we should look for a semantics for the attitudes that is specifically sensitive to a shifting domain of ordered worlds.

3.2 Some attitude semantics

Kadmon & Landman argue that we should understand the semantics of *glad* and *sorry* as being intimately linked to the semantics of *want*. Very roughly—we shall have reason to become more sophisticated—the idea is that *glad that p* is almost the same as *want that p*, and that *sorry that p* is almost the same as *want that not p*. Now, if *want* is UE, then *glad* will be too and *sorry* will be DE, which is what we want. Therefore, let's start with investigating the semantics of *want*, about which we can find much detailed discussion in Heim (1992).

A Kratzer-style semantics for *want*

A rather simple-minded semantics for *want* may look like this:

(36) $[\![wants]\!]^{f,g}(p)\,(\alpha)\,(w) =$ True iff $\forall w' \in max_{g(\alpha,w)}(f(\alpha, w))\colon w' \in p$
    'Among the worlds $f(\alpha, w)$, the ones that maximally correspond to α's preferences in w are all p-worlds.'

This semantics is built on the two ingredients already mentioned: a set of worlds among which the subject α is said to have preferences and a way of ordering these worlds according to α's preferences in the evaluation world:

(37) (i) The 'modal base function' f is a function from pairs of an individual and a world to a set of worlds.
    (ii) The 'ordering source function' is a function from pairs of an individual and a world to a set of propositions (in the case of *want* to a set of propositions forming the subject's preferences).

In the semantics of the attitude predicate, the ordering source (a set of propositions) is used to induce an ordering on the worlds in the modal base. The attitude predicate then makes a claim about those worlds in the modal base that maximally satisfy the preferences given by the ordering source. Here is how a set of propositions induces an ordering on a set of worlds:

(38) For any set of propositions P, we define a strict partial order $<_P$:
    $\forall w', w''\colon (w' <_P w''$ iff $\forall p \in P(w'' \in p \rightarrow w' \in p)$ and
    $\exists p \in P(w' \in p \,\&\, w'' \notin p))$
    w' is better than w'' according to P iff all propositions in P that hold in w'' also hold in w' but some hold in w' that do not also hold in w''.

Then we pick the worlds in the modal base that are best by the ordering $<_p$:

(39) For a given strict partial order $<_p$ on worlds, define the selection function maxP that selects the set of $<_p$-best worlds from any set X:
$$\forall x \subseteq W: \max_P(X) = \{w \in X: \neg \exists w' \in X: w' <_p w\}.$$

The semantics of attitude predicates will be sensitive to these two parameters, and in turn the lexical semantics of an attitude can specify which kind of parameters it needs to work with. Since there are plenty of different kinds of attitudes, there must be plenty of different kinds of modal bases and ordering sources. It is therefore not quite right to specify the semantics as in (36) with just a particular f and g. Perhaps it would be best to give attitude predicates an index and think of the parameters f and g as being functions from indices to modal base functions and ordering source functions. We revise (36) to this:

(40) $[\![wants_i]\!]^{f,g}(p)(\alpha)(w) = \text{True iff } \forall w' \in \max_{g_i(\alpha,w)}(f_i(\alpha, w)): w' \in p$

As it stands, this is still not complete. The predicate *want* will carry conditions on which kinds of modal base and ordering source need to be supplied to it. We'll turn to the question of what modal base it wants in a moment. But for now, note that by squinting at (40) one can see that it will make *want* an upward entailing operator (which is what we need to carry out the Kadmon & Landman project). If all of the maximally preferred worlds in $f(\alpha, w)$ are p-worlds and every p-world is a q-world, then all maximally preferred worlds in $f(\alpha, w)$ will be q-worlds. If p entails q, then *John wants p* should entail *John wants q*.

What is the right modal base for *want?*

A simple idea is that $f(\alpha, w)$ for *want* is the set of worlds compatible with α's beliefs in w, the set of doxastically accessible worlds for $\Rightarrow$ in w.

(41) $[\![wants_i]\!]^{f,g}(p)(\alpha)(w)$ is defined only if $f_i(\alpha, w) = DOX(\alpha, w)$, where $DOX(\alpha, w)$ is the set of worlds compatible with α's beliefs in w. If defined, $[\![wants_i]\!]^{f,g}(p)(\alpha)(w) = \text{True iff}$
$\forall w' \in \max_{g_i(\alpha,w)}(f_i(\alpha, w)): w' \in p$

This idea has some nice immediate consequences: it predicts that worlds which run counter to the subject's beliefs will be irrelevant for evaluating *want*-sentences. Consider in this light an example from Heim:

(42) I want to teach Tuesdays and Thursdays next semester.

Heim writes about this: 'Suppose this sentence is intuitively true as spoken by me today. Is it therefore the case ... that I teach Tuesdays and Thursdays next semester in all the worlds that are compatible with everything I desire? No. In worlds that are compatible with everything I desire I actually don't teach at all' (Heim 1992: 195). If a *want*-predication only makes a claim about the relative desirability of the worlds compatible with the subject's beliefs, we can explain (42). If the speaker assumes that she will teach on some days next semester, all (42) claims is that among those worlds (in all of which she does teach) the most desirable ones are ones where she teaches Tuesdays and Thursdays.

Unfortunately, this analysis makes not so nice predictions as well. If α believes that p, then we would predict directly that α wants that p. If all of α's belief worlds are p-worlds then the most desirable ones among those will also be p-worlds. This is not good. For one, it would predict that since the speaker of (42) believes she will teach next semester, it would be true to say of her that she wants to teach next semester. That does not seem right. Similar problems were noted by Stalnaker:

Suppose I am sick. I want to get well. But getting well entails having been sick, and I do not want to have been sick. Suppose there was a murder. I want to know who committed the murder. But my knowing who committed the murder entails that the murder was committed, and I never wanted the murder to have been committed (Stalnaker 1984: 89).

Heim discusses a possible solution to this problem. Suppose that α *wants that p* presupposes that α neither believes that p nor believes that not p. We may try something like this:

(43) $[\![wants_i]\!]^{f,g}(p)\,(\alpha)\,(w)$ is defined only if
    (i) $f_i(\alpha, w) = \mathrm{DOX}(\alpha, w)$
    (ii) $f_i(\alpha, w) \cap p \neq \varnothing$
    (iii) $f_i(\alpha, w) - p \neq \varnothing$
    If defined, $[\![wants_i]\!]^{f,g}(p)\,(\alpha)\,(w) = $ True iff
    $\forall w' \in \max_{g_i(\alpha, w)}(f_i(\alpha, w)): w' \in p$

You can only want something of whose fact you are not (yet) convinced. This would not make it false of the speaker of (42) that she wants to teach next semester, but it would make the claim suffer from presupposition failure. Heim suggests we can live with that result.

But Heim argues that we need to refine the analysis yet some more. It seems that you can actually want something which you firmly believe will happen, as long as you think it's up to you whether it will happen. Heim's example is this:

(44) (John hired a babysitter because) he wants to go to the movies tonight.

According to Heim, (44) 'certainly does not suggest in any way that John has the slightest doubt about where he will be tonight'. So, Heim suggest that instead of setting $f(\alpha, w)$ to the set of doxastically accessible worlds it should be taken to be the set of worlds compatible with 'everything that α in w believes to be the case no matter how he or she chooses to act'. This will always be a superset of the set of worlds compatible with everything α in w believes simpliciter. Let's call this new set $DOX^*(\alpha, w)$.

So far so good, then. It seems that with the proper understanding of what modal base *want*-claims quantify over, we have a satisfactory analysis of *want* that makes it an upward entailing operator. Here for future reference is the analysis we have arrived at:

(45) $[\![wants_i]\!]^{f,g}(p)$ (α) (w) is defined only if
  (i) $f_i(\alpha, w) = DOX^*(\alpha, w)$
  (ii) $f_i(\alpha, w) \cap p \neq \emptyset$
  (iii) $f_i(\alpha, w) - p \neq \emptyset$
  If defined, $[\![wants_i]\!]^{f,g}(p)$ (α) (w) = True iff
  $\forall w' \in max_{g_i(\alpha,w)}(f_i(\alpha, w)): w' \in p$

This semantics describes how the meaning of *want* (through its definedness conditions) requires a certain kind of domain of quantification. It leaves open the nature of the ordering source, which is a gap that will have to be filled. For our purposes it suffices to say that *want* will take as its ordering source a set of propositions specifying the preferences of α in w.

The apparent non-monotonicity of *want*

There is a kind of case, introduced by Asher (1987), that may suggest that we are wrong in analyzing *want* as an UE operator. Asher says that the following inference is invalid:

(46) Nicholas wants to get a free trip on the Concorde.
  So, Nicholas wants to get a trip on the Concorde.

He writes: 'If I want to ride on the Concorde and not pay for it, it doesn't necessarily follow that I also want to ride on the Concorde *simpliciter*. It may mean bankruptcy!' (Asher 1987: 171).

One reaction to such examples, which is the one found in Heim (1992), is to see them as reasons to give up the monotonic analysis. What may be going on here is this: Nicholas thinks that among the worlds compatible

with his beliefs some are more likely than others. It is not likely that he will get a free trip, more likely he will have to pay $3000. What *Nicholas wants a free trip* means is that the most likely worlds in which he gets a free trip are better than the most likely ones in which he doesn't. This can be true, while at the same time it is false that *Nicholas wants a trip*, because the most likely worlds in which he takes a trip are expensive worlds and thus worse than the most likely worlds in which he doesn't take a trip at all.

Heim presents such a non-monotonic analysis (involving both a desirability ordering and a likelihood or similarity ordering). Here's a version:[10]

(47) If defined,
$$[wants_i]^{f,g}(p) \ (\alpha) \ (w) = \text{True iff } \forall w' \in f_i(\alpha, w): \text{Sim}_{w'}(p) <_{g_i(\alpha w)} \text{Sim}_{w'}(\neg p)$$
where $\text{Sim}_{w'}(p) = \{w' \in p: w' \text{ resembles } w \text{ no less than any other world in } p\}$

This is, of course, much more complex than what we had before. The claim is not anymore that within the modal base (roughly given by the subject's beliefs) the worlds that maximally conform to the subject's preferences are *p*-worlds. Now, we would say this: for any world w' in the modal base (still roughly given by the subject's beliefs in *w*) the *p*-worlds most similar to w' are better (according to the subject's preferences in *w*) than the *non p*-worlds most similar to w'.[11]

If this analysis is right, *want* is not UE. This does not leave much hope for the claim that *regret* is DE. Can it be that Asher and Heim are wrong? Well, I am not convinced that the Asher inference in (46) is really invalid.

What is clearly invalid is the following: *Nicholas wants a free trip on the Concorde. Therefore, no matter how much it costs Nicholas wants a trip on the*

---

[10] This formulation is essentially found in (31) on p. 193 of Heim's paper.

[11] There is another possibility which doesn't use an additional ordering but makes *want*-claims much stronger and thereby also destroys UE inferences:

(i) $[wants_i]^{f,g}(p) \ (\alpha) \ (w) = \text{True iff } f_i(\alpha, w) \cap p <_{g_i(\alpha, w)} f_i(\alpha, w) - p$

This would not merely require that the best relevant worlds are *p*-worlds, but that all relevant *p*-worlds are better than any of the relevant *non p*-worlds. This would also take care of the Asher case. Assume that Nicholas wants a free trip, which under this semantics means that all the free trip scenarios are more desirable than any of the scenarios where Nicholas either doesn't go on the trip or pays through the nose for it. One cannot infer that all of the scenarios where Nicholas takes a trip, whether for free or for a lot of money, are better than any of the scenarios where he doesn't go on a trip.

However, I think that the semantics in (i) is far too strong. It would make *I want a Honda Civic* equivalent to something like *I want any Honda Civic that I can think of.* To me, for *I want a Honda Civic* to be true it is enough if there is a small number of Honda Civics (perhaps even just one) with which I would be maximally happy. So, (i) cannot be right.

Concorde. But, why should we read the conclusion *Nicholas wants a trip on the Concorde* as being equivalent to *No matter how much it costs Nicholas wants a trip on the Concorde?* It may well be that Asher reads the conclusion of (46) as having such a strong meaning. At least, I am not sure what else he might mean by his paraphrase *I want to ride on the Concorde simpliciter.*

Heim (class notes) has a variant on Asher's example, which makes matters clearer:

(48) John is in a furniture store, looking at a couch that has a very scary price-tag. The salesman comes up to him and the following conversation takes place:
Salesman: Would you like to buy this couch?
John: No.
Salesman: Would you like to buy it at a 25% discount?
John: Yes.

What are we entitled to conclude? At the time of John's first answer, it would seem that one could truthfully say: *John does not want to buy this couch.* At the time of John's second answer, one could truthfully say: *John wants to buy this couch at a 25% discount.* But it is not clear that we have to conclude that at the very same time in the very same context the following statements are true:

(49) (i) John doesn't want to buy this couch but he wants to buy this couch at a 25% discount.
  (ii) John wants to buy this couch at a 25% discount but he doesn't want to buy this couch.

In fact, these seem hopelessly contradictory.

It seems that John's first statement in (48) has to be understood against the background of a set of worlds in which the couch has exactly the price stated on the price tag. Among those, the most desirable worlds are not ones where John buys the couch. By the time of John's second statement, more worlds are made accessible: apparently the couch can be bought at a 25% discount. Worlds where John does buy the couch at that discount are highly desirable, so John wants to buy the couch at that price. Is the first statement still true in the new situation? I don't think so: it is now false that John doesn't want to buy the couch. In the new context, we would have to say that John doesn't want to buy the couch at its original price. But that is not the same (anymore) as saying that John doesn't want to buy the couch.

Note that with the proposal in (45), we can say more precisely what is going on. At first, John's set of worlds DOX*(John, w) will only contain worlds where the couch costs as much as is stated on the price tag that he sees. Then the salesman asks the question whether John wants to buy the

couch at a 25% discount. But because of the semantics of *want* in (45), this question carries the presupposition that there are worlds in DOX*(John, w) where the couch is bought at a 25% discount. This is not the case for the old DOX*(John, w), so John will have to revise his opinions about what is possible. DOX*(John, w) will be widened to include worlds where the couch is bought at a 25% discount. Within this newly widened modal base, John's preferences (which have not changed) will single out the worlds where he does buy the couch at that discount as the most desirable worlds

So, perhaps a UE analysis of *want* is possible after all, as long as we pay attention to the shifting grounds of context. If *p* entails *q*, α *wants p* will entail α *wants q*, as long as it is assumed that *q* has the required relation to the set of worlds DOX*(α, w). *q* must be an open issue with respect to what α believes will happen no matter how s/he chooses to act. If, as we have done, these conditions are formalized as presuppositions of *want*-statements, then we have that α *wants p* Strawson-entails α *wants q*. In a context where *q* does not satisfy these conditions, the assertion of α *wants q* may nevertheless be appropriate, as long as the context can be adjusted so as to then satisfy the conditions. This is what happens in the couch example. But inferences are checked with respect to a constant context, and so this kind of example does not provide counter-evidence against the claim that *want* is Strawson-UE.[12]

We can now move on to the factive attitudes. We need to establish, using the tools developed so far, that *sorry* is DE in the relevant sense and that *glad* is not.

3.3 'Glad' and 'sorry'

As I mentioned, Kadmon & Landman argue that there is a very special connection between *glad* and *sorry* (and their cousins) on the one hand and *want* on the other. They write

It is deeply rooted in the meaning of *sorry* that if you are sorry about a fact A, given a certain perspective, that is because you want the opposite, i.e., ¬A, relative to the same perspective ... Vice versa, if the facts are opposite (the negation) of a certain wish you have, then you must be sorry about them, at least given the same perspective relative to which you have your wish. Similarly, there is a special connection between being glad and one's wishes. If you are glad that A, that is because you want A, relative to the same perspective. Vice versa; if A is a fact, and you want A, then you must be glad that A, given the same perspective.

[12] Shifting contextual grounds and in particular shifts in which possible worlds are relevantly accessible for the evaluation of attitude sentences and other intensional sentences are discussed in a seminal paper by David Lewis (1979), see also his 1996 paper on knowledge claims.

There are two immediate problems with this discussion. It can't really be *want* that is involved in the semantics of *glad* and *sorry*. We saw that a *want*-claim is really only appropriate *vis-à-vis* matters that are still open in a certain sense. *Glad* and *sorry*, on the other hand, are, of course, *factive* predicates. They are attitudes towards matters that are now settled. What must be involved in their semantics is a notion of desire that is not restricted to open matters. In fact, Heim suggests that the factive attitudes like *glad* and *sorry* involve counterfactuality. This prima facie astounding claim ('factive verbs involve counterfactuality!?') can be supported as follows. Both *α is glad that p* and *α is sorry that p* presuppose that *α* believes that *p* (and often indicate that *p* is in fact true). The attitude that is expressed by such sentences compare the world as *α* believes it to be to the world as *α* believes *it would be if p were not true*. The element of counterfactuality is therefore forced to be present by the factivity of these predicates.

The other immediate worry about Kadmon & Landman's discussion is that they often talk as if attitudes are relations to *facts*, whereas most semantic treatments see them as *propositional* attitudes. Some of the persuasiveness of their argumentation may come from the particular way in which they talk about facts. I will come back to this point in a little while.

Glad

Now, let's talk about *glad*. Based on our prior discussion of *want*, we can perhaps start with the following idea:

(50) $[\![glad_i]\!]^{f,g}(p)$ (α) (w) is defined only if
    (i) $DOX(\alpha, w) \subseteq p$
    (ii) $DOX(\alpha, w) \subseteq f_i(\alpha, w)$
    (iii) $f_i(\alpha, w) = \cap p \neq \varnothing$
    (iv) $f_i(\alpha, w) - p \neq \varnothing$
    If defined, $[\![glad_i]\!]^{f,g}(p)$ (α) (w) = True iff
    $\forall w' \in max_{g_i(\alpha, w)}(f_i(\alpha, w))$: $w' \in p$

The first definedness condition (i) here implements the factivity of *glad*. One can only be glad that *p* if one thinks that *p* is true. We have a *want*-style truth-condition: all the most desirable worlds among the relevant worlds are *p*-worlds. For this to make sense, the set of worlds $f(\alpha, w)$, among which *α*'s preference are reported, must not be equated with $DOX(\alpha, w)$. Otherwise, you would be glad about anything that you believe to be true. So, again we impose a presupposition that among the worlds in $f(\alpha, w)$, there are some *p*-worlds and some *non p*-worlds, just as we did for *want* in (45). These conditions in (iii) and (iv) have the immediate effect that

some counterfactual worlds (some *non p*-worlds) must be in the modal base, must be relevant alternatives. To prevent the modal base from totally going off into counterfactual space, we require it to properly contain the subject's doxastic alternatives (condition (ii)).

How is the modal base selected? How many 'counterfactual worlds' are relevant? The definedness conditions (i)–(iv) in (50) provide a lower bound on the extent of the modal base of worlds compared by a *glad*-statement. Can we say more about upper bounds? I can think of two plausible ways one might go about selecting a proper modal base for a *glad*-statement. (i) Start with $DOX(\alpha, w)$, add to it those *non p*-worlds that are most similar to w, then add any worlds not already in $DOX(\alpha, w)$ that are more similar to w than the most similar *non p*-worlds. The *glad*-statement would then amount to the claim that the worlds that best fit the subject's preferences are *p*-worlds even if one takes into consideration those *non p*-worlds that are most similar to the evaluation world. (ii) Many if not all evaluative attitude sentences involve attitudes towards the outcome of actions or events in the past. Perhaps to figure out which worlds are relevant to a *glad that p*-statement, we should go back in time to a point where it was not yet determined that *p* would happen or turn out true. From that point on, go forward and collect all worlds that could have developed out of that situation, most importantly any world that might have resulted from 'deciding' that *not p*. For all I know, this second procedure is not importantly distinct from the first. Perhaps, what would be used in finding those *non p*-worlds that are most similar to w is exactly this method of going back into the past to a point where it was not yet sure that *p*. I will leave this aspect of the semantics of attitudes under-developed. I can dimly see an elaboration of the analysis that dynamically updates the domain of quantification for attitude ascriptions in the manner just sketched. But exploring this would lead us too far astray here. We will concentrate on the question of the monotonicity behavior of the attitude predicates.

Kadmon & Landman are concerned with showing that *glad* is not DE, even in a sophisticated context-dependent semantics (they are answering a concern raised by Linebarger). So, they add the following discussion:

If I'm glad he bought a car, then it clearly agrees with my wishes that he bought a car. What ought I to feel, then, about his buying a Honda? Well, the purchase of the Honda has certainly satisfied my wish that he buy a car. Hence, I can easily be glad that he bought it, qua car, i.e. in as much as it satisfies the wish. This explains the fact that speakers often judge that we have here DE on a constant perspective. However, note that I am not necessarily FORCED to be glad that he bought the Honda, because buying a Honda is not by any means required for satisfying my wish. My wish could be satisfied in another way, for example, by buying a Toyota . . . Hence, *glad* is not DE.

While I think Kadmon & Landman are right when they say that *glad* is not DE, I think that they nevertheless have the wrong semantics for *glad*. What is wrong about it is the following inference, which they implicitly endorse in the preceding quote:

(51) α wants p

  p

  (α knows p)

  $\therefore$ α is glad that p

I think that's invalid. I want to buy a Honda Civic, because a Honda Civic would maximally satisfy my needs for a car. I buy a Honda Civic, which turns out to be a real lemon. It is still true that I desire a Honda Civic, just not the one I actually have. So, I'm not glad I bought a Honda Civic. I wanted to, I did, I still think it was a good idea, but I'm not glad I did.

In more abstract terms: according to (50), in order to be glad that *p* it is enough that the most desirable relevant worlds are *p*-worlds and that the actual world (or at least any world compatible with the subject's beliefs) is a *p*-world, but it is not necessary that the actual world is one of those *p*-worlds which are among the most desirable worlds. This seems insufficient to me. Kadmon & Landman may actually agree with me, since they like to talk about facts rather than propositions. What they say is that if A is a fact and I want A, then I must be glad about A. For facts that may in fact be right (even though I don't know what an analysis running on facts would look like).

I would like to suggest that the semantics in (50) be replaced with the following semantics, which is still an upward entailing analysis:[13]

(52) If defined, $[\![glad_i]\!]^{f,g}(p)\,(\alpha)\,(w)$ = True iff

  $DOX(\alpha, w) <_{g_i(\alpha,w)} (f_i(\alpha, w) - p)$

The new analysis demands that among the compared worlds the ones that are compatible with the subject's beliefs (which are all *p*-worlds and which for all the subject believes could be the actual world) are better (according to the subject's preferences in the actual world) than all the relevant

---

[13] One could consider another way of addressing my concern, a non-monotonic analysis with the following truth-condition (everything else stays the same):

  (i) $f_i(\alpha, w) \cap p <_{g_i(\alpha,w)} f_i(\alpha, w) - p$

This analysis demands that not just the actual world (or rather any world that for all the subject believes could be the actual world) but all relevant *p*-worlds be more desirable than any relevant *non p*-world. This is stronger than my proposal in (52). I think that my proposal is correct. Assume again that I want to buy a Honda Civic, because certain Honda Civics would maximally satisfy my needs for a car. I end up buying one of those Honda Civics. I will be glad that I bought a Honda Civic, even if there are relevant (but counterfactual) worlds where I buy an awful Honda Civic which are less desirable than worlds where I buy an adequate Toyota Corolla.

non p-worlds in the modal base. Note that the main result that Kadmon & Landman care about still holds: *glad* is not DE, in fact it is UE. For me to be glad that I bought a Honda, the actual world (or rather any world that for all I believe may be the actual world) will have to be better than any relevant world in which I don't buy a Honda (including worlds where I buy a Toyota and worlds where I don't buy a car at all). Hence, I will be glad that I bought a car.

That's enough about *glad* for the time being. We will soon come back to how one should deal with apparently coherent sequences such as *I'm not glad that I bought a car, but at least I'm glad that I bought a Honda Civic* or *I'm not glad she died but I'm glad she died peacefully/in her sleep/with a clear conscience* (example due to the anonymous reviewer).

Sorry

We turn to *sorry* (and *regret*, and with suitable adjustments *surprised, amazed*, etc.). We need to find a semantics that is Strawson-DE and that will allow us to deal with the apparent counter-examples like (31). There are some obvious candidates to consider. Two that I will discuss here are the 'mirrors' of the two meanings we explored for *glad*. Here they are (we assume the same definedness conditions as for *glad*):

(53) If defined, $[\![sorry_i]\!]^{f,g}(p)\,(\alpha)\,(w) = $ True iff
$\forall w' \in \max_{g_i(\alpha,w)}(f_i(\alpha, w)): w' \notin p$

(54) If defined, $[\![sorry_i]\!]^{f,g}(p)\,(\alpha)\,(w) = $ True iff
$(f_i(\alpha, w) - p) <_{g_i(\alpha,w)} \mathrm{DOX}(\alpha, w)$

The meaning in (53) demands that in the best of all relevant worlds, the complement *p* (which is presupposed to be true according to the subject's beliefs) is not true. The meaning in (54) requires that all relevant *non p*-worlds have to be better than the subject's belief worlds. The first meaning is Strawson-DE, the second is not.[14]

---

[14] One can, of course, imagine other analyses. In fact, with a good glass of red wine and paper and pencil it is astonishingly easy to come up with candidate analyses that are not blatantly implausible. Wouldn't it be nice if the language learner got some obvious clues about which meanings are serious contenders (such as the fact that NPIs are licensed by *sorry*, hence the meaning should be Strawson-DE)?

Two further candidates one might consider (again, with the rest of the schema understood):

(i) $\exists w' \in (f_i(\alpha, w) - p): w' <_{g_i(\alpha,w)} \mathrm{DOX}(\alpha, w)$

(ii) $f_i(\alpha, w) - p <_{g_i(\alpha,w)} f_i(\alpha, w) \cap p$

The very weak meaning in (i) can be dismissed fairly easily. It would make us people with very many regrets. For me to regret that *p* all that would be needed is that *some non p*-world is more desirable than the actual *p*-world (or rather any world that I believe might be the actual world). I would regret anything that is such that if it were not true, I *might* be better off. That doesn't seem right. To regret something you have to be convinced that you would in fact be better off without it.

Kadmon & Landman argue that *sorry* is downward entailing, which would mean we should pick (53) as the correct analysis. Here's their argument. They consider three examples and argue as follows:

(55) I'm glad he bought a car.
(56) I'm sorry he bought a car.
(57) He bought a Honda.

Now consider (56). If I'm sorry he bought a car, I clearly wish he had bought no car. What ought I to feel, then, about his buying a Honda? I ought to be sorry about it, qua car. In fact, and this is where (56) differs from (55), I MUST be sorry about it, qua car. That is because refraining from buying a Honda is an absolute requirement for satisfying my wish. I cannot prefer for my wish to be satisfied in 'another way'. Hence, *sorry* is DE (on a constant perspective). The crucial difference should be clear: the wish associated with (55) places no particular demand on the truth value of (57), whereas the wish associated with (56) requires it to be false.

As far as I understand, this argument goes as follows: *I'm sorry that he bought a car* entails *I wish that he had not bought a car*, which in turn entails *I wish that he had not bought a Honda*, which in turn entails *I'm sorry that he bought a Honda*. Thus by transitivity, *I'm sorry that he bought a car* entails *I'm sorry that he bought a Honda*. QED

The argument thus relies on two assumptions. One is that *sorry that p* is equivalent to *wish that not p*. The other is that *wish* is upward entailing. (Note by the way that Kadmon & Landman have switched from relating *sorry* to *want* to the more adequate relationship between *sorry* and *wish*).

If we adopt the following semantics for *wish* and pair it off with the meaning for *sorry* in (53), we have what this argument needs:

(58) $[\![wish_i]\!]^{f,g}(p)(\alpha)(w)$ is defined only if
   (i) $\text{DOX}(\alpha, w) \cap p = \varnothing$
   (ii) $\text{DOX}(\alpha, w) \subseteq f_i(\alpha, w)$
   (iii) $f_i(\alpha, w) \cap p \neq \varnothing$
   (iv) $f_i(\alpha, w) - p \neq \varnothing$
   If defined, $[\![wish_i]\!]^{f,g}(p)(\alpha)(w) = \text{True}$ iff
   $\forall w' \in \max_{g_i(\alpha,w)}(f_i(\alpha, w)): w' \in p$

The meaning in (ii) is the mirror of the meaning for *glad* mentioned in fn. 13. It demands that all relevant *non p*-worlds be better than any relevant *p*-world, not just the ones that the subject believes might be the actual world. Since this is the mirror of the rejected meaning for *glad*, it must be rejected for the same reason. According to this meaning, to be sorry that I bought a Honda, non-Honda worlds don't just have to be better than the actual Honda-world but better than any Honda I might have bought within the realm of additional counterfactual worlds in the modal base. That is too strong. It's enough that I loathe the Honda I actually bought.

According to this analysis, *wish* is a counterfactual attitude: by (i), *p* is presupposed to be false in the subject's belief worlds. But the relevant modal base also contains *p*-worlds. The quantificational claim is that among the relevant worlds the ones that are best according to the subject's preferences are all *p*-worlds. (I leave it as an exercise for the reader to establish that this semantics indeed makes *wish that not p* and *sorry that p* equivalent).

Curveball #1: Iatridou's conditional semantics for *wish*[15]

While there is nothing obviously implausible about Kadmon & Landman's idea that *sorry that p* and *wish that not p* are equivalent, there is one reason to think that the semantics of *wish* is a little more involved (whether that would carry over to *sorry* is another issue). Iatridou (1998) describes how *wish*-statements in English and even more so their correspondents in other languages share properties of counterfactual conditionals. Roughly, English *wish that p* is often expressed in other languages as *would want that p*. Iatridou shows that the verbal mood/tense/aspect morphology found in the complement *p* is identical to that found in the antecedents of counterfactual conditionals, while the morphology found on predicate *would want* (lexicalized in English as *wish*) is identical to that found in the consequent of counterfactual conditionals. She proposes the following representation:[16]

(59) α wishes that p: α thinks that if p were the case, s/he would be glad that p.

This representation is supposed to be a paraphrase that captures the meaning of *wish that p* as it would be delivered by an explicit compositional semantics. There is no implied claim that at some abstract syntactic level *wish*-sentences have the structure of (59).

Here are some noteworthy properties of this analysis. First, note that the counterfactual conditional is embedded under *think*. Iatridou does this to explain why the counterfactual presupposition (that *p* is not true) does not (automatically) project out of the attitude context. *Wish*-statements presuppose that the subject of the wish believes the complement to be contrary-to-fact; they don't necessarily presuppose that the speaker shares that belief. In my proposal in (58), this 'subject-orientation' of the

---

[15] For people not familiar with baseball: a curveball is a ball thrown in a particularly unexpected and confounding way designed to catch the hitter off guard. I use the term here in one of its colloquial meanings, roughly a problem coming at one from an unexpected direction.

[16] I have adjusted Iatridou's formulation somewhat, but the difference is not relevant here.

counterfactual presupposition is directly encoded in condition (i). A simpler paraphrase (*If p were true, α would be glad that p*) would not capture this fact.

The second noteworthy feature of the paraphrase in (59) is the fact that *p* occurs twice in it. I think that this move is necessary for two related reasons. Treating *p* as just a counterfactual antecedent in the meaning of *wish* would not really capture the meaning of the construction. *α wishes that p* means more than *if p were the case, α would be happy*. The simple conditional would be true if the subject would be happy no matter whether *p* or *not p*. But that of course is not enough for the *wish*-claim to be true. Not only does the subject have to be happy but she has to prefer *p* to *not p*.

The double role of *p* is also beneficial in that it prevents *wish* from displaying the same monotonicity behavior as counterfactual conditionals. As we will see in section 4, there is good reason to think that conditional antecedents are DE environments. But then an analysis of *α wishes that p* as meaning *if p were the case, α would be happy* would predict that the complement of *wish* is also a DE environment. That is of course ludicrous. If anything, it should be a UE environment (at least that is what Kadmon & Landman and I assume). Iatridou's paraphrase makes *wish* non-monotonic and thus correctly predicts that it will not license NPIs in its complement.[17]

In Iatridou's proposal, *α wishes that p* is equated with *α thinks that if p were the case, s/he would be glad that p*. Under natural assumptions about belief-ascriptions, counterfactual conditionals, and the semantics of *glad* (as we analyzed it earlier), we would have the claim that for any world w'

[17] Let me mention on the side that Iatridou's proposal is reminiscent of Pesetsky's (1991: ch. 10) ideas about the analysis of complement *if*-clauses in English, as in the following:
    (i)   Emma would like it if Orestes were here.
Pesetsky builds on prior work by Williams (1974: especially 157–64), Steriade (1981), and Pullum (1987). Examples similar to (i) include also:
    (ii)   Emma would be happy if Orestes were here. ·   (under one of its two readings)
         Emma would have liked Orestes to be here.
Karina Wilkinson (pc to Pullum) had noticed that these *if*-clauses cannot contain NPIs, in contrast to normal conditional antecedents:
    (iii) *I would like it if anyone were to ask me about the painting.   (Pesetsky's (235), p. 178)
         If anyone were to ask me about the painting, I would like it.
Pesetsky explains the absence of NPIs in complement *if*-clauses by saying that there is a Post-LF rule that turns the *if*-clause into a conditional antecedent but simultaneously copies its 'non-quantificational content' (anything other than the ingredient that makes it into a conditional antecedent) into the complement position of the attitude predicate.
    (iv) Emma would like it if Orestes were here.
       ~> If Orestes were here, Emma would like it that Orestes is here.
NPIs are of course not licensed in positive attitude complements:
    (v) *Emma likes it that anyone came to visit her.
So, again it is the double role of these complements that prevents them from licensing NPIs. For the cases discussed by Pesetsky, this is a pressing concern since it is an *if*-clause that fails to license NPIs (while it is a fact that 'normal' *if*-clauses do license NPIs). For Iatridou's case, there may be a more direct solution: maybe *wish* doesn't involve a counterfactual antecedent at all.

compatible with what α believes in the actual world (and keep in mind that α must believe that p is not true), all of the p-worlds w″ most similar to w′ are such that w″ is better (according to α's preferences) than any relevant *non p*-world.[18]

Iatridou's discussion is very interesting, but if she is correct it would disturb the delicate balance of my edifice. It makes *wish that p* non-monotonic, which threatens the downward monotonicity of *sorry that p*, at least under the assumption that Kadmon & Landman were right in claiming the two constructions to be equivalent. I could of course give up that assumption, but I can't see any good direct arguments against it. The other course of action is to hope that the morphosemantic facts that Iatridou wants to explain can be captured without making *wish*-sentences into covert counterfactual conditionals. But that can only remain a hope for now.

3.4 *Shifting contexts*

After these detailed ruminations about the semantics of various attitude predicates (a discussion which seems so interesting that one has to be glad that it became necessary in our ongoing quest for a theory of NPI licensing), we will now tackle some of the examples that are prima facie problematic for the monotonic analyses. These will have to be analyzed as involving shifting contexts.

The following sequences are both coherent:

(60) Sandy is glad that Robin bought a car, but Sandy is sorry/not glad that Robin bought a Honda.

(61) Sandy is sorry that Robin bought a car, but Sandy is glad/not sorry that Robin bought a Honda.

Using the monotonic analyses for *glad* and *sorry* that we developed in the previous section, the coherence of (60) is straightforward. Assume that we have an ordering of worlds where the best worlds are ones where Robin buys a Toyota, in the next best worlds Robin buys a Honda, and in the worst worlds Robin doesn't buy any car. Then, one should be glad that Robin bought a car (because any world where she doesn't is worse than the actual world; even though she could have done better by choosing a different car, her choice was good enough to beat worlds without a car). And one should be sorry about her buying a Honda, because in the best worlds she buys a Toyota.

---

[18] A similar (also non-monotonic) meaning for *wish* is sketched by Heim (1992: 205): 'John wishes you were gone means "John thinks that if you *were* gone he *would* be in a more desirable world than he is in *because* you are not gone"'.

The coherence of (61) is not so straightforward. To be sorry that Robin bought a car means that the most desirable relevant worlds are non-car worlds, and thus *a fortiori* non-Honda worlds. Then it can't be that the actual Honda-world is better than *any* non-Honda world. What we have to say is that there is a shift in which worlds are being considered. The idea is that for the evaluation of the second sentence only worlds in which Robin actually buys a car are compared. Then, all the second sentence would claim is that the actual Honda-world is better than any of the worlds in which Robin buys *a car other than a Honda*. That is a claim which is compatible with the claim expressed by the first sentence with respect to a larger set of worlds. Linebarger (1987: 369) in fact had the pertinent intuition when she observed that such sequences involve 'covert conditionalization'. She noted: 'John may regret that he assaulted a fellow patron in a restaurant because he was arrested and fined, but it may also be true that *given that he did so*, he is glad that it was the obnoxious George Smith that he assaulted.'

The claim then is that the coherence of (61) does not point to the necessity of a non-monotonic semantics for *sorry* (etc.) but is merely a reflection of the shiftable nature of context. Specifically, from one constituent sentence to the other the modal base function has changed. As natural as this is, the usual conception of validity of inferences is formulated to check truth-preservation with respect to a *constant* context. And in the view taken here, (61) does not involve a constant context and so its coherence does not furnish an argument against the monotonicity of *sorry* with respect to a constant context.

Direct arguments against non-monotonic analysis?

In the case of *want*, we had reasons independent of NPI licensing to reject the non-monotonic account. We argued that problematic cases depend on a change in the modal base, in particular a widening of the possibilities considered. We saw that it is quite incoherent to attribute the relevant attitudes 'in one breath' as it were:

(49) (i) !!John doesn't want to buy this couch but he wants to buy this couch at a 25% discount.
(ii) !!John wants to buy this couch at a 25% discount but he doesn't want to buy this couch.

Can we produce similarly convincing examples in the case of *sorry*? Reversing the sequence does seem to introduce some oddness:

(62) ?Sandy is glad that Robin bought a Honda
but Sandy is sorry that Robin bought a car.

According to our story, the narrower attitude needs to be understood as implicitly conditionalized. In the good example in (61), the first sentence furnishes the proposition that Robin bought a car, which can then be used as the background for the second sentence. In (62), this natural sequence is disturbed.

It also seems to me that focus structure plays a role. (61) is naturally read as having narrow focus on *Honda*, evoking contrasts such as *Toyota, Mercedes*, etc. And Linebarger's example above has a cleft: *He is glad that it was the obnoxious George Smith that he assaulted.* So, the focus structure of the second sentence may support an interpretation of the attitude against the back-ground of an understood conditionalization. If these supporting clues are removed, we do not seem to get sequences that are quite as coherent:

(63) Bill and Mary came ⇒ Bill came.
!! I'm glad that Bill and Mary came. But I'm sorry that Bill came.
!! I'm sorry that Bill came. But I'm glad that Bill and Mary came.
Better:
I'm sorry that Bill came. But I'm glad he came with Mary.

The effect of focus can also be contravened by using the expression *qua* which is quite often employed by Kadmon & Landman.

(64) !!Sandy is sorry that Robin bought a car, but Sandy is glad/not sorry that Robin bought a Honda qua car.

I guess the idea would be that *Sandy is glad Robin bought a Honda qua car* signals that the gladness is not one that concerns the choice of a Honda over possible alternative brands.[19]

We then do have some reasons to think that the coherence of (61) does not follow as a matter of course from a non-monotonic semantics for *sorry* but depends on contextually signaled narrowing of the domain of quantification of the monotonic operator *sorry*.

A closer look at the contribution of focus

We actually need to consider carefully the contribution of focus to the interpretation of our attitude statements. If focus induces a change in the modal base, all is to the good: changes in the modal base disqualify the relevant patterns as potential counter-examples to downward inferences *in a constant context.* But if the contribution of focus is more internal to the semantics of the attitude ascriptions, for example if focus is taken as an internal restriction on the very same modal base that the broader

---

[19] It would be intriguing to study the semantics and pragmatics of *qua*. I think it might also be illuminating to see what it does to the interpretation of adjectives:

(i) John is large, qua man, but quite small, qua basketball player.

statement was sensitive to, focus-sensitivity would make the semantics non-monotonic. Let's think about this.

That there is an effect of focus structure on the interpretation of attitude ascriptions is quite well known.[20] This phenomenon of 'association with focus' is one that attitude ascriptions share with many other constructions (especially quantificational ones). The now widely accepted story of how this phenomenon works involves a semantics of focus based on *alternatives*. Let us assume that for any expression α, we compute not only its usual denotation $[\![\alpha]\!]$ but also its focus denotation $[\![\alpha]\!]_F$, which will be a set of denotations which are focus-alternatives to $[\![\alpha]\!]$. The principal effect of focus is to introduce a set of alternatives to the focused item. This can then be passed on 'up the tree' and can lead to sets of alternatives for bigger expressions. The focus on *Honda* in (61) first evokes a set of relevant contrasts X to Honda. Higher up what we get are alternative propositions of the form 'Robin bought a X'. For our purposes, what will suffice is reference to the union of this set of propositions. What we would have as the relevant object is the set of worlds in which Robin buys a car of one of the relevant brands.

So, by the time we have interpreted the complement *Robin bought a [Honda Civic]*$_F$, we have as its ordinary interpretation the proposition that Robin bought a Honda Civic and as its focus interpretation the proposition that Robin bought a car. Within modern theories of focus semantics, there are now two main lines of research about what to do with these two propositions: (i) focus structure induces a presupposition about the context (Rooth 1992, 1996), (ii) focus structure can be accessed directly by the semantics of other operators (Krifka 1991a).

The presuppositional theory will fit in nicely with my analysis of why affective attitudes license NPIs. The narrow focus in the complement would merely signal the presence in the context of the proposition that Robin bought a car; a condition that is clearly satisfied. That this proposition is also taken as an implicit restriction on what worlds are relevant to the second attitude ascription in (61) would not be attributed directly to the focus semantics. Focus would just be a defeasible signal.

The theory that employs reference to focus alternatives directly in the semantics of focus-sensitive operators would however clash with my theory. Here is how such a semantics may look like:

(65) *Focus is used directly by the semantics*

$[\![sorry_i\, thatp]\!]^{f,g}(\alpha)\,(w)$ is defined only if

(i) $\mathrm{DOX}(\alpha, w) \subseteq [\![p]\!]^{f,g}$

(ii) $\mathrm{DOX}(\alpha, w) \subseteq f_i(\alpha, w)$

(iii) $f_i(\alpha, w) \cap [\![p]\!]^{f,g} \neq \varnothing$

---

[20] The first author to draw attention to the phenomenon was Dretske (1972, 1975).

(iv) $f_i(\alpha, w) - [\![p]\!]^{f,g} \neq \varnothing$

If defined, $[\![sorry; thatp]\!]^{f,g}(\alpha)(w) = \text{True}$

iff $\forall w' \in \max_{g_i(\alpha, w)}(f_i(\alpha, w)) \cap \cup [\![p]\!]_F^{f,g})w' \in [\![p]\!]^{f,g}$

Under the second proposal, there is no change in the contextual parameters. Instead, internal to the semantics of *sorry*, we directly employ the focus denotation of the complement sentence by intersecting the contextually supplied set of relevant worlds with the set of worlds evoked by the focus structure of the complement. If we adopt this analysis, the example in (61) will not involve context shift and thus would be a crucial counter-example against the thesis that *sorry* is DE.

To maintain our story, then, we have to come down on the side of the purely presuppositional theory of focus structure (further discussion of the attraction of the indirect theory of association with focus can be found in von Fintel 1994). Examples where focus effects a narrowing of the set of worlds compared by an attitude predicate are to be seen as examples where a contextual parameter is shifted. Whether a predicate is DE can only be judged in contexts where no such context shift occurs.

Curveball #2: the DE-ness of focus-sensitive *only*

Danny Fox (pc) asks what happens when we try to determine whether *only* is Strawson-DE, once we move beyond the simple cases of *only* plus proper name that were discussed in section 2. Consider:

(66) There **only** was any precipitation in [MEDford]$_F$.
(67) There **only** was precipitation in [MEDford]$_F$.
(There was rain in Medford.)

∴ There **only** was rain in [MEDford]$_F$.

We can take (66) as a representative of a much larger class of cases. *Only* is a cross-categorial operator that (like negation and conjunction) can combine with many kinds of expressions. As discussed in von Fintel (1997), a popular story (developed by Rooth 1985) is to reduce all such uses of *only* to a base case where *only* acts as a propositional modifier, a kind of generalized negation with respect to an understood set of alternative propositions. (66) would claim that among the relevant alternative propositions the only true one is that there was precipitation in Medford. The focus structure of the sister proposition of *only* (which I call its 'prejacent', following medieval semantic terminology) gives clues about the set of relevant alternatives. With the focus on *Medford* in (66), we have a signal that only propositions

about the weather in places that are relevant alternatives to Medford count as negated by *only*.

As we see in (66), these uses of *only* also license NPIs in their prejacent. And as we see in (67), we also observe Strawson-DEness. So, there does not seem to be any problem. There is one worry, however. We need to make sure that the semantics of *only* is stated in a way that delivers these facts correctly. And here, it is easy to go astray. In von Fintel (1997), for example, essentially the following two entries for propositional *only* are given as options between which one may choose freely:

(68) For all sets of relevant alternative propositions C, propositions p, r, worlds w:

    a. $[\![only]\!]$ (C) (p) (w) is defined only if p(w) = True

    If defined, $[\![only]\!]$ (C) (p) (w) = True iff $\forall r \in C$: r(w) = True $\Rightarrow$ (p $\Rightarrow$ r)

    b. $[\![only]\!]$ (C) (p) (w) is defined only if p(w) = True and (ii) $\neg \exists r \in C$: p $\Rightarrow$ r

    If defined, $[\![only]\!]$ (C) (p) (w) = True iff $\forall r \in C$: r(w) = True $\Rightarrow$ (p = r)

What I was dealing with at that point was the problem of how to make sense of the claim that the only true proposition in the set of relevant alternatives is the one that it rained in Medford. Clearly, we don't want this to make the impossible claim that none of the entailments of that proposition are true. So, one possibility (the one in (a)) is the claim that *only* makes is weakened to the claim that the only true propositions in the set of relevant alternatives are the proposition that it rained in Medford and any of the propositions entailed by that one. The other possibility, the one in (b), is to say that propositions entailed by the prejacent are not legitimate alternatives, that only sets of propositions C will be accepted that do not contain any entailments of the prejacent. At the time, I did not see any empirical reason to choose one option over the other, and I proceeded to adopt (b) for concreteness.

This was exactly the wrong choice as we can see now. Adopting (b), we get a machinery that predicts that *only* will not license downward inferences, within the same set of alternatives C (because C is devoid of any propositions entailed by the prejacent). The semantics in (a) correctly gives *only* the required Strawson-DE property.

In this story as well, we need to say that the focus structure of the prejacent does not *force* a different set of alternatives C on us. If for the premise in (67) only propositions of the form 'there was precipitation in X' are allowed in C, and for the conclusion only propositions of the form 'there was rain in X' are allowed in C, there would have to be two different

sets C of alternatives, and thus no constant context within which to assess the downward inference.

Much of these complications would be avoided if we treated cases like (66) as not involving propositional *only* but a logical form where the associate of *only* (here, the proper name *Medford*) forms a constituent with *only*. The logical structure would be: *Only Medford is such that there was rain in it*. Then, the simpler semantics discussed in section 2 would be all we need. But this move is one I cannot consider any further here.

3.5 *Where we are*

We have reached a point where we are exploring the idea that the notion of downward entailment that NPI licensing is sensitive to has two important properties: (i) we need Strawson Entailment, because presuppositions carried by the conclusion in downward inferences don't seem to disrupt NPI licensing; (ii) contextual parameters need to be kept constant *even if in a natural conversation they would normally evolve in a certain way*. With these tools in hand, we turn now to the semantics of conditionals.

4 ANTECEDENTS OF CONDITIONALS

We will go through the same kind of dialectic as before. NPIs are licensed where there is reason to think downward entailingness does not hold. A context-savvy semantics may get around that. But first a red herring needs to be taken care of.

4.1 *The problem*

NPIs are licensed in the antecedent of conditionals:[21]

(70) a. If John subscribes to <u>any</u> newspaper, he is **probably** well informed.
  c. If he has <u>ever</u> told a lie, he **must** go to confession.
  d. If you had left <u>any</u> later, you **would** have missed the plane.

This fact is problematic from the point of view of the Fauconnier–Ladusaw generalization, since conditional antecedents are not obviously downward

---

[21] Partee (1992) shows that the licensing of NPIs in *if*-clauses is not some dumb mistake of the grammar. If the *if*-clause restricts a non-universal quantifier, where it is uncontroversial that there is no downward monotonicity, NPIs are not allowed:

(i) #Sometimes, if a man feeds a dog any bones, it bites him.

monotone contexts. In fact, in the modern semantic and philosophical literature on conditionals, it is now taken for granted that conditionals are not monotonic in their antecedent, in that they are claimed to not validate the inference pattern known as Strengthening the Antecedent. Some spectacular failures of this pattern are as follows:

(70) a. If <u>I strike this match</u>, it **will** light.
   ⇏ If <u>I dip this match into water and strike it</u>, it **will** light.
   b. If <u>John stole the earrings</u>, he **must** go to jail.
   ⇏ If <u>John stole the earrings and then shot himself</u>, he **must** go to jail.
   c. If <u>kangaroos had no tails</u>, they **would** topple over.
   ⇏ If <u>kangaroos had no tails but used crutches</u>, they **would** topple over.

Heim (1984) proposed to weaken Ladusaw's analysis to only demand a 'limited' kind of downward monotonicity. The idea was to keep more of the environment constant. Instead of checking for entailment between two arbitrary antecedents, let them only differ in the place of the NPI. Kadmon & Landman (1993) show that this move is not enough. The failure of Strengthening the Antecedent extends to cases where only the position of the crucial NP is manipulated:

(71) If <u>John subscribes to a newspaper</u>, he **must** be well informed.
   ⇏ If <u>John subscribes to a newspaper that he can't read</u>, he **must** be well informed.

Kadmon & Landman suggest that conditionals are downward monotone after all, as long as we keep the context constant for the whole stretch of the argument. The same idea is advocated by Katz (1991). None of these authors addresses the fact that the standard Stalnaker-Lewis analysis (and Kratzer's variations, which those authors primarily refer to) actually does claim to keep the value of contextually supplied parameters constant.

The standard non-monotonic semantics for conditionals

Let us see how the standard non-monotonic analysis of conditionals might go. We will formulate it in terms by now familiar to us. Conditionals are seen by Kratzer as the result of combining a modal operator with a restrictive *if*-clause, which is analyzed as narrowing down the modal base that the modal operator is sensitive to. The ordering source is used to select a particular subset from the set of worlds in the modal base that the

antecedent is true in. The idea is that different conditional constructions differ in what exactly their modal base and their ordering source is.

For example, counterfactuals (involving the modal *would*) take as their modal base the entire set of possible worlds and as their ordering source a set of propositions that is 'totally realistic' (taken together these propositions uniquely describe the evaluation world). Such an ordering source will thus single out from the worlds in the modal base those that are most similar to the evaluation world, most similar according to the particular description of the evaluation world encoded in the ordering source. In Lewis (1981), it is shown that this treatment of counterfactuals makes exactly the same predictions about the logical behavior of counterfactuals as the familiar Stalnaker–Lewis analyses (modulo possible refinements).

Other conditionals may take as their modal base a set of propositions encoding the current state of knowledge in the utterance situation and use as their ordering source assumptions about the stereotypical course of events. This would make them 'epistemic conditionals'. And so on.

The general schema for the interpretation of conditionals according to this view can be detected in this proposal for the semantics of the universal modal *would* as restricted by an *if*-clause:

(72) $[\![\textit{if } p, \textit{ would}_i q]\!]^{f,g}(w) = \text{True iff}$
$\forall w' \in \max_{g_i(w)}(f_i(w) \cap [\![p]\!]^{f,g}): w' \in [\![q]\!]^{f,g}$

Such a sentence claims that among the worlds in the modal base (assigned to the evaluation world) in which the antecedent is true, the ones that are best according to the relevant ordering source are all worlds in which the consequent is true.[22]

---

[22] Formulating the Stalnaker-Lewis semantics in terms of selecting the maximally best worlds is only possible under what Lewis calls the Limit Assumption, which Lewis in fact rejects. Stalnaker, on the other hand, defends the assumption against Lewis' arguments by saying that in actual practice, in actual natural language semantics and in actual modal/conditional reasoning, the assumption is eminently reasonable. Kratzer is persuaded by Lewis' evidence and does not make the Limit Assumption; hence her semantics for modals is more convoluted than what I present here. I side with Stalnaker, not the least because it makes life easier. For discussion, see Lewis (1973) and Stalnaker (1984: ch. 7, esp. 140–2). Further arguments *against* the Limit Assumption can be found in Herzberger (1979) and Pollock (1976: 18–20). Further arguments *for* the Limit Assumption can be found in Warmbrod (1982).

An assumption that I do not share with Stalnaker is the Uniqueness Assumption: that for any *p*, the set of maximally best *p*-worlds in the modal base is in fact a single *p*-world. Stalnaker uses this assumption to justify the Conditional Excluded Middle, the inference from *not (if p, would q)* to *if p, would not q*. To deflect worries that no realistic context will supply an ordering source (similarity measure in the case of counterfactuals) that is sharp enough to distinguish *p*-worlds so finely as to single out one *p*-world as the best one, Stalnaker sketches an alternative employing supervaluation. My own preferred approach to the Excluded Middle is to say that those modal operators that validate this inference carry a presupposition of homogeneity—they presuppose that all *p*-worlds behave uniformly with respect to the truth of *q*. This approach is sketched in von Fintel (1997).

Consider then:

(73) If I had struck this match, it would have lit.
    $\not\Rightarrow$ If I had dipped this match into water and struck it, it would have lit.

According to the non-monotonic account, this inference is semantically invalid. The premise merely claims that the most highly ranked worlds in which I strike this match are such that it lights. No claim is made about the most highly ranked worlds in which I first dip this match into water and then strike it. The reason for the invalidity of (73) becomes perceptible because it is natural to assume that the most highly ranked worlds in which I strike this match are ones where I make sure that the match is dry.

Quite importantly, the account maintains that the crucial examples are cases where the context remains relevantly the same throughout the examples. Lewis attempts to demonstrate this with the following kind of example:

(74) If the USA threw its weapons into the sea tomorrow, there would be war; but if the USA and the other nuclear powers all threw their weapons into the sea tomorrow, there would be peace.

This speaker simultaneously asserts a counterfactual conditional and the negation of a counterfactual conditional derived from it by Strengthening the Antecedent. Lewis deliberately put this example in the form of a single run-on sentence, with the counterfactuals conjoined by semicolons and *but*. This is meant to ensure that the context stays constant throughout.

There are two options within our current project of trying to salvage the Fauconnier–Ladusaw account of NPI licensing: (i) We could try to argue against the standard non-monotonic account of conditionals. Kadmon & Landman and Katz wish to pursue the first line of analysis; but as mentioned above, they mistakenly assume that Kratzer's semantics is a monotonic analysis and so they don't actually go very far. (ii) We could try to make something out of the fact that the non-monotonic semantics involves a 'superlative' ingredient (this strategy was suggested to me by Angelika Kratzer). Let us pursue the second option for a moment.

4.2 Superlatives and NPI licensing

Superlatives license NPIs:

(75) Emma is the **tallest** girl to <u>ever</u> win the dance contest.

But again, we are disappointed to find that superlatives are not straight-forwardly DE:

(76) Emma is the **tallest** girl in her class.
 $\not\Rightarrow$ Emma is the **tallest** girl in her class to have learned the alphabet.

Superlatives are, however, Strawson-DE:

(77) Emma has learned the alphabet.
 Emma is the tallest girl in her class.
 $\Rightarrow$ Emma is the tallest girl in her class to have learned the alphabet.

This behavior is clearly visible only when the superlative predicate is used in an otherwise simple sentence. When a superlative is used inside an argument definite description, a DE inference will not go through.

(78) The largest mammals are over 100 feet long. (Angelika Kratzer, pc)
 The largest mice are over 100 feet long.

What is going on? The only discussion of NPI licensing in superlatives that I am aware of is Hoeksema (1986b), where he notes that superlatives have a limited kind of downward monotonicity, an idea that seems very similar to the idea of Strawson Entailment. Let me try to spell out how a story might go.

We need a meaning for the superlative morpheme that fits into the structure α *is the P-est* Q. Assume that the definite determiner here is idle or is interpreted as part of a unit together with the superlative morpheme. Observe that α *is the P-est* Q presupposes that α *is a* Q. It then asserts that among the other Qs all are P to a smaller degree than α.

(79) $[\![$the $\dots$ -est$]\!]$ (P) (Q) (α) is defined only if Q (α) = True
 If defined, $[\![$the $\dots$ -est$]\!]$ (P) (Q) (α) = True iff
 $\forall x \neq \alpha$: (Q(x) = True $\rightarrow$ ιd P (x) (d) < ιd P (α) (d))

For this to work Q needs to be a normal one-place predicate like *girl in her class* or *girl in her class who has learned the alphabet*.[23] The predicate that the superlative attaches to must be a gradable one, which I treat here as a relation between an individual x and a degree d.[24]

By this semantics, it is invalid to infer downward in the position of Q, since the conclusion may suffer from presupposition failure. But the inference *is* Strawson-valid, since under the assumption that α satisfies

---

[23] I will not discuss the interesting fact that relative clauses under a superlative operator are often realized as *infinitival* relative clauses.

[24] There is, of course, relevant literature on comparatives and gradable predicates and so on, which I entirely ignore here. My modest aim is to present an analysis that is just plausible enough for our purposes.

the subset property, the truth of the superlative premise guarantees the truth of the superlative conclusion.

Now, (79) treats the superlative as a predicate-creating operator taking both a comparison predicate (P) and a domain predicate (Q) as its arguments. When the superlative predicate is used as the predicate inside a definite description, for example when *the tallest girl* is used as the subject of a sentence, the Strawson-DEness of the superlative will be obliterated. Consider for example:

(80) The tallest girl in Emma's school is over four feet tall.
    $\not\Rightarrow$ The tallest girl in Emma's class is over four feet tall.

The inference in (80) is not even Strawson-valid. If the conclusion of (80) has a presupposition it is that there is a girl in Emma's class that is taller than any other girl in Emma's class. This presupposition will not rescue the inference in (80). What would be needed is the assumption that the tallest girl in Emma's class is the same as the tallest girl in Emma's school. But this assumption is not available as a presupposition.

The reason why the superlative licenses NPIs is then that the superlative morpheme as defined in (79) is Strawson-DE. The fact that once such a structure is further embedded (as it is when it used as the restriction of a definite description operator), downward inferences are not licensed does not change the *local* Strawson-DEness of the superlative morpheme. This property only 'shines through' in an unadulterated way when the superlative predicate is used predicatively, without being embedded in further structure.

Now what about conditionals?

My explanation for NPI licensing by superlatives depended on teasing apart the contribution of the superlative morpheme from that of the higher environment (including the definite description operator). This is of course legitimate because we can observe the superlative morpheme at work in an unadulterated fashion in predicative uses, as in (75). To carry this kind of story over to conditionals, one would have to claim that at a level relevant to the determination of NPI licensing, we are dealing not with simple modal operators like *would* but with complex *structures* of the form *the closest P-worlds*. In other words, what is a paraphrase in the usual semantics for conditionals must be turned into a hypothesis about the relevant *syntactic structure* of conditionals.

This is a project that I find implausible.

4.3 *A Monotonic semantics for conditionals?*

Our remaining hope then is that conditionals are in fact monotonic and that the examples that are supposed to show the invalidity of Strengthening the Antecedent involve context shifts.

Authors who claim that this would be the correct analysis are not too hard to find. Approaches that give a monotonic semantics and explain non-monotonic behavior by appeals to pragmatics are very popular for indicative conditionals. Accounts in a roughly Gricean vein (Grice 1967, 1989) continue to be refined, see especially Jackson's work (1979, 1984, 1987, 1990). Other pragmatically informed analyses include Veltman (1986) and McCawley (1993: 548ff). Stalnaker himself defines a notion of 'reasonable inference' in his paper on indicative conditionals (Stalnaker 1975) which bears some resemblance to what I will develop here. Pragmatic approaches to the interpretation of subjunctive conditionals are advocated in some form or another by Warmbrod (1981a, b, 1983), Wright (1983), and Lowe (1990, 1995).

None of these works quite goes where I go in my paper 'Counterfactuals in a Dynamic Context' (von Fintel 2000). They employ pragmatic considerations to *explain away* counterexamples. Instead, I admit that the counterexamples are genuine but deny that they force us to adopt a static non-monotonic semantics. Rather, the source of non-monotonicity is in the dynamics of domains of quantification. I refer the interested reader to that paper for details. The idea of my dynamic analysis is that there is an ever-widening 'modal horizon', which we will model by a function from worlds to sets of worlds, i.e. an accessibility function. We will use the ordering source parameter for a slightly different purpose than in Kratzer's system: it will be used to make sure that the evolution of the modal horizon is such that it always forms a well-behaved 'Lewis-sphere' around the evaluation world. The domain of worlds that a modal/conditional operator quantifies over is given by intersecting its antecedent with the modal horizon.

(81) *Admissible Modal Horizons*
    A function D from worlds to sets of worlds is an admissible modal horizon with respect to the ordering source g iff
    for any world w, $\forall w''(w'' \leq_{g(w)} w' \rightarrow w'' \in D(w))$.

(82) $[\![would_i]\!]^{D,g}$ (if p) (q) (w) is defined only if
    (i) D_i is admissible with respect to g_i
    (ii) $D_i(w) \cap p \neq \emptyset$ (p is compatible with the modal horizon)

(83) If defined, $[\![would_i]\!]^{D,g}$ (if p) (q) (w) = True iff
    $\forall w' \in D_i(w) \cap p$: q(w) = True.

Counterfactual sentences carry a presupposition about the context: the context has to be well behaved in the sense of providing an admissible modal horizon. The counterfactual is further only defined for worlds to which the modal horizon assigns a set of accessible worlds that is compatible with the proposition expressed by the antecedent. (In effect, counterfactuals thus carry an existence presupposition with respect to their domain of quantification.)

In cases where the context does not already provide a modal horizon that obeys the compatibility condition, a new context will have to be created to satisfy the presupposition of the counterfactual. A natural procedure for repairing the context would of course be to minimally expand the modal horizon so as to assign antecedent worlds to any evaluation world. In my paper 'Counterfactuals in a Dynamic Context', this procedure of enlarging the modal horizon in a constrained way is formalized in a fully dynamic semantics. Here, we can stick with the static presuppositional format.

The semantics given here is Strawson-DE. *Under the assumption* that the modal horizon is wide enough to be compatible with both p and p & r, the inference from *if p, would q* to *if p* & *r, would q* will be valid. The same inference is not valid without making this assumption. Thus, if p & r is not compatible with the initially selected modal horizon, the presupposition of the conclusion would not be satisfied. A constrained change of the context would then occur, selecting a slightly larger modal horizon. But this may in fact remove the justification for holding the initial conditional.

Evidence for context shift

There is some evidence that the counter-examples to SA in fact involve such context shifts. Consider the clear contrast between Lewis' example and a variant due to Irene Heim:

(74) If the USA threw its weapons into the sea tomorrow, there would be war; but if the USA and the other nuclear powers all threw their weapons into the sea tomorrow, there would be peace.

(84) ??If all nuclear powers threw their weapons into the sea tomorrow, there would be peace; but if the USA threw its weapons into the sea tomorrow, there would be war.

In (84), the two counterfactuals claimed to be consistent by Lewis are reversed in their order and the sequence does not work as before. The reason seems intuitively clear: once we consider as contextually relevant worlds where all nuclear powers abandon their weapons, we can't ignore them when considering what would happen if the USA disarmed itself. We

seem to be in need of an account that keeps track of what possibilities have been considered and doesn't allow succeeding counterfactuals to ignore those possibilities. An account according to which the context remains constant throughout these examples would not expect a contrast between these two orders.

As mentioned before, Lewis takes the coherence of (74) as a sign that the context does not shift. Similarly, Edgington (1995: 252f.) presents the following scenario: 'a piece of masonry falls from the cornice of a building, narrowly missing a worker. The foreman says: "If you had been standing a foot to the left, you would have been killed; but if you had (also) been wearing your hard hat, you would have been alright".' Edgington says, quite correctly, that the building foreman's remarks constitute 'a single, pointful piece of discourse'. One can easily read them as a shrewd way of putting the suggestion that the worker should wear her hard hat at all times.

The fact that (74) and Edgington's example are 'single pointful pieces of discourse' argues against attempts at dismissing them as cases of illicit equivocation. But there is no argument here against the idea that the context can and does change over the course of simple pointful discourses. The proper diagnosis would seem to be that over the course of (74) the modal horizon properly expands, but that over the course of (84) it cannot shrink. This asymmetry is unexpected if one maintains there is no context change.

Note also that if someone utters (74), someone else can then rejoin that the initial conditional is 'no longer' true:[25]

(85) A: If the USA threw its weapons into the sea tomorrow, there would be war; but if the USA and the other nuclear powers all threw their weapons into the sea tomorrow, there would be peace.

B: But that means that if the USA threw its weapons into the sea tomorrow, there wouldn't NECESSARILY be war.[26]

---

[25] What I mean by 'no longer true' is not that the objective facts have changed. It is the parameters of the discourse that have changed so that the proposition expressed by the first counterfactual in the initial context can no longer be expressed by the same linguistic expression in the new context. Compare the fact that the claim that *France is hexagonal* may be true in a context where it is preceded by *Italy has the shape of a boot*, but cease to be true in a later context where the standards of precision have been sharpened.

[26] Note that the stress on *necessarily* is required. B cannot say (i) or (ii):

(i) But that means that if the USA threw its weapons into the sea tomorrow, there would NOT be war.

(ii) But that means that it is not TRUE that if the USA threw its weapons into the sea tomorrow, there would be war.

The reason for this is investigated in von Fintel (1997). The idea is that bare conditionals obey the Excluded Middle and that therefore negating them either has a very strong meaning or needs to be done by using an explicit operator that does not obey the Excluded Middle.

B′: But that means that if the USA threw its weapons into the sea tomorrow, there might NOT be war.

This is unexpected under the standard static approach. If we go back to the simpler antecedent, the domain of quantification should shrink back to the closest worlds where just the USA disarms, ignoring the far-fetched worlds where all nuclear powers become meek. But that doesn't seem to happen.[27]

The Strawson-validity of strengthening the antecedent

According to the analysis given in (81)–(83), Strengthening the Antecedent is not a truth-preserving inference; it will not be classically valid. This is a prediction that is confirmed by the counter-examples to the pattern that we saw. Moving to a strengthened antecedent may lead to failure of the presupposition that the current modal horizon is compatible with the antecedent. So, from a premise that is true with respect to a particular context, we might move to a conclusion that puts conditions on the context that are not satisfied. The immediate result would be presupposition failure. But in the normal case, the context may be adjusted so as to assume a slightly wider modal horizon. But then the truth of the antecedent would have to be reassessed with respect to the new context.

Nevertheless, SA is Strawson-valid in my system. This means that under the additional assumption that the presuppositions of the conclusion are satisfied, the downward inference will be truth-preserving. The additional assumption is essentially one that ensures that the antecedent is not so novel or bizarre as to fall outside the current modal horizon. If the modal horizon is already wide enough to accommodate the strengthened antecedent, SA will be safe.

If this is the correct analysis for conditionals, the Fauconnier–Ladusaw analysis of NPI licensing can be maintained. Conversely, if the Fauconnier–Ladusaw analysis of NPI licensing is attractive, we have reason to adopt the Strawson-monotonic semantics for conditionals.

5 CONCLUSION

We have explored the prospects of the Fauconnier–Ladusaw approach to NPI licensing. With judicious choices in the semantics of particularly problematic constructions and with specific assumptions about what kind

---

[27] Note that B′ seems to rely on an inference from *if p and r, would q* to *if p, might q*. This pattern is invalid in the standard system, but is valid in mine.

of entailment notion we are after, one can in fact stick to the Fauconnier–Ladusaw idea. The project is, of course, only as attractive as the particular semantic analyses that it depends on.

The most striking aspect of the project is that it requires checking of inferences only in cases where the contextual parameters are not affected by the assertion of the test sentences. In as much as this is not a particularly common situation, especially from the point of view of dynamic semantics, the account, if true, would be quite surprising.

We must also note that the algebraic condition on NPI licensing can only be a necessary condition. As shown by Linebarger, there are effects that cannot easily be accounted for on a pure Fauconnier–Ladusaw-style analysis. Linebarger (1991) discusses conditional antecedents as licensing environments for NPIs and concludes:

These two environments—conditionals and relative clauses headed by universal quantifiers—represent particularly good cases of licensing by downward entailingness, and particularly unconvincing cases of licensing by NI [= negative implicatures (KvF)]. They are unconvincing because they license NPIs with no 'conscious' negative implicature, although NPIs are unacceptable in environments where the possibility of NOT P is absent (p. 178).

Linebarger herself thus admits that conditionals do provide prima facie support for the Fauconnier–Ladusaw analysis over hers (although she does not discuss the problem of the apparent failure of Strengthening the Antecedent, save for a reference to Heim's paper). The one interfering factor that she mentions can be illustrated as follows:

(86) a. #If you drink any water, you'll feel a whole lot better.
     b. #If you think Bob had any fun, you should have seen Fred!

As noted by Lakoff (1969), NPIs are unacceptable in conditionals used as promises as in (a) or in conditionals in which the possibility that the antecedent is false is remote as in (b). What exactly the actual nature of the additional conditions on NPI licensing are must remain an open question here.

Acknowledgements

first presented informally to the MIT LF-Reading Group, after which Danny Fox asked a question which led to the material in the final subsection of section 3.4. The paper was then presented at the second annual conference on *Sinn und Bedeutung* in Berlin (December 1997). I distinctly recall helpful and encouraging remarks from Graham Katz, Rob van Rooy, and Arnim von Stechow, which they may, of course, deny in retrospect. Thanks also to an anonymous reviewer for the *Journal of Semantics*.

I wish to dedicate this paper to the memory of Jim McCawley.

KAI VON FINTEL
Department of Linguistics and Philosophy
Massachusetts Institute of Technology
E39-245, 77 Massachusetts Avenue
Cambridge, MA 02139, USA
e-mail: fintel@mit.edu
http://web.mit.edu/fintel/

REFERENCES

Asher, Nicholas (1987), 'A typology for attitude verbs and their anaphoric properties', *Linguistics and Philosophy*, 10, 127–89.

Atlas, Jay (1993), 'The importance of being only: testing the neo-Gricean versus neo-Entailment paradigms', *Journal of Semantics*, 10, 301–18.

Atlas, Jay (1996), '"Only" noun phrases, pseudo-negative generalized quantifiers, negative polarity items, and monotonicity', *Journal of Semantics*, 13, 265–328.

Corcoran, John (1973), 'Meanings of implication', *Diálogos*, 25, 59–76.

Dretske, Fred (1972), 'Contrastive statements', *Philosophical Review*, 81, 411–37.

Dretske, Fred (1975), 'The Content of Knowledge', in Bruce Freed, A. Marras, & P. Maynard (eds), *Forms of Representation*, North-Holland, Amsterdam, 77–93.

Edgington, Dorothy (1995), 'On conditionals', *Mind*, 104, 235–329.

Fauconnier, Giles (1975), 'Polarity and the scale principle', *Chicago Linguistics Society*, 11, 188–99.

Fauconnier, Giles (1979), 'Implication reversal in a natural language', in Franz

Guenthner & S. J. Schmidt (eds), *Formal Semantics and Pragmatics for Natural Languages*, Reidel, Dordrecht, 289–302.

von Fintel, Kai (1994), 'Restrictions on quantifier domains', Ph.D. dissertation, Graduate Student Linguistics Association (GLSA), University of Massachusetts, Amherst.

von Fintel, Kai (1997), 'Bare plurals, bare conditionals, and *only*', *Journal of Semantics*, 14, 1–56.

von Fintel, Kai (2000), 'Counterfactuals in a dynamic context', to appear in Michael Keystowicz (ed.), *Ken Hale: A Life in Language*, Cambridge, MA: MIT Press.

Giannikidou, Anastasia (1997), 'The landscape of polarity items', Ph.D. dissertation, Rijksuniversiteit Groningen.

Grice, Paul (1967), 'Logic and conversation', MS.

Grice, Paul (1989), *Studies in the Way of Words*, Harvard University Press, Cambridge, MA.

Heim, Irene (1984), 'A note on polarity sensitivity and downward entailingness', *NELS*, 14, 98–107.

Heim, Irene (1992), 'Presupposition projection and the semantics of attitude verbs', *Journal of Semantics*, 9, 183–221.

Herzberger, Hans (1979), 'Counterfactuals

and consistency', *Journal of Philosophy*, 76, 83–8.

Hoeksema, Jack (1983), 'Negative polarity and the comparative', *Natural Language and Linguistic Theory*, 1, 403–34.

Hoeksema, Jack (1986a), 'Monotonicity phenomena in natural language', *Linguistic Analysis*, 16, 25–40.

Hoeksema, Jack (1986b), 'Monotonie en Superlatieven', in C. Hoppenbrouwers (ed.), *Proeven van Taalwetenschaap*, Tabu, Groningen, 38–49.

Horn, Larry (1969), 'A presuppositional approach to *only* and *even*', *Proceedings of the Chicago Linguistic Society (CLS)*, 5, 98–107.

Horn, Larry (1979), '*Only, even*, and conventional implicature', MS (LSA Paper, Los Angeles).

Horn, Larry (1996), 'Exclusive company: *only* and the dynamics of vertical inference', *Journal of Semantics*, 13, 1–40.

Horn, Larry (1997), 'All John's children are as bald as the King of France: existential import and the geometry of opposition', *Proceedings of the Chicago Linguistic Society (CLS)*, 33.

Iatridou, Sabine (1998), 'The grammatical ingredients of counterfactuality', MS, MIT.

Israel, Michael (1996), 'Polarity sensitivity as lexical semantics', *Linguistics and Philosophy*, 19, 6, 619–66.

Jackson, Eric (1995), 'Weak and strong negative polarity items: licensing and intervention', *Linguistic Analysis*, 25, 181–208.

Jackson, Frank (1979), 'On assertion and indicative conditionals', *Philosophical Review*, 88, 565–89.

Jackson, Frank (1984), 'Two theories of indicative conditionals', *Australasian Journal of Philosophy*, 62, 67–76.

Jackson, Frank (1987), *Conditionals*, Blackwell, Oxford.

Jackson, Frank (1990), 'Classifying conditionals', *Analysis*, 50, 2, 134–47.

Kadmon, Nirit & Landman, Fred (1993),

'*Any*', *Linguistics and Philosophy*, 16, 353–422.

Katz, Graham (1991), 'The downward entailingness of conditionals and adversatives', *FLSM*, 2, 217–43.

Klima, Edward (1964), 'Negation in English', in Jerry Fodor & J. J. Katz (eds), *The Structure of Language*, Prentice-Hall, Englewood Cliffs, 246–323.

Kratzer, Angelika (1981), 'The notional category of modality', in H. J. Eikmeyer & H. Rieser (eds), *Words, Worlds, and Contexts: New Approaches in Word Semantics*, de Gruyter, Berlin, 38–74.

Krifka, Manfred (1990), 'Polarity phenomena and alternative semantics', in Martin Stokhof & L. Torenvliet (eds), *Proceedings of the Seventh Amsterdam Colloquium*, ITLI, Amsterdam, 277–302.

Krifka, Manfred (1991a), 'A compositional semantics for multiple focus constructions', *SALT* 1, 127–58.

Krifka, Manfred (1991b), 'Some remarks on polarity items', in Dietmar Zaefferer (ed.), *Semantic Universals and Universal Semantics*, Foris, Berlin, 150–89.

Krifka, Manfred (1994), 'The semantics and pragmatics of weak and strong polarity items in assertions', *Semantics and Linguistic Theory (SALT)* 4, 195–219.

Krifka, Manfred (1995), 'The semantics and pragmatics of polarity items', *Linguistic Analysis*, 25, 209–57.

Ladusaw, William (1979), 'Polarity sensitivity as inherent scope relations', Ph.D. dissertation, University of Texas at Austin.

Ladusaw, William (1980a), 'Affective *or*, factive verbs, and negative polarity items', *Chicago Linguistics Society*, 16.

Ladusaw, William (1980b), 'On the notion "affective" in the analysis of negative polarity items, *Journal of Linguistic Research*, 1, 1–23.

Lakoff, Robin (1969), 'Some reasons why there can't be any *some–any* rule', *Language*, 45.

Lewis, David (1973), *Counterfactuals*, Blackwell, Oxford.

Lewis, David (1979), 'Scorekeeping in a language game', *Journal of Philosophical Logic*, 8, 339–59.

Lewis, David (1981), 'Ordering semantics and premise semantics for counterfactuals', *Journal of Philosophic Logic*, 10, 217–34.

Lewis, David (1996), 'Elusive knowledge', *Australasian Journal of Philosophy*, 74, 549–79.

Linebarger, Marcia (1980), 'The grammar of negative polarity', Ph.D. dissertation, MIT.

Linebarger, Marcia (1987), 'Negative polarity and grammatical representation', *Linguistics and Philosophy*, 10, 325–87.

Linebarger, Marcia (1991), 'Negative polarity as linguistic evidence', *Chicago Linguistic Society*, 27, 2, 165–88.

Lowe, E. J. (1990), 'Conditionals, context, and transitivity', *Analysis*, 50, 2, 80–7.

Lowe, E. J. (1995), 'The truth about counterfactuals', *Philosophical Quarterly*, 45, 41–59.

McCawley, James (1993), *Everything that Linguists Have Always Wanted to Know about Logic but Were Ashamed to Ask*, Chicago University Press, Chicago, (second edn, much revised).

Partee, Barbara (1992), 'On the "scope of negation" and polarity sensitivity', in Eva Hajicova (ed.), *Functional Description of Language: Proceedings of the Conference, Prague, 24–27 November 1992*. Faculty of Mathematics and Physics, Charles University, Prague.

Pesetsky, David (1991), 'Zero syntax, vol. 2', MS, MIT.

Pollock, John (1976), *Subjunctive Reasoning*, Reidel, Dordrecht.

Pullum, Geoffrey (1987), 'Implications of English extraposed irrealis clauses', *ESCOL*, 260–70.

Rooth, Mats (1985), 'Association with focus', Ph.D. dissertation, University of Massachusetts at Amherst.

Rooth, Mats (1992), 'A theory of focus interpretation', *Natural Language Semantics*, 1, 75–116.

Rooth, Mats (1996), 'Focus', in Shalom Lappin (ed.), *The Handbook of Contemporary Semantic Theory*, Blackwell, Oxford, 271–97.

Stalnaker, Robert (1975), 'Indicative conditionals', *Philosophia*, 5, 269–86.

Stalnaker, Robert (1984), *Inquiry*, MIT Press, Cambridge, MA.

Steriade, Donca (1981), 'Complement *if*, MS, MIT.

Strawson, P. F. (1952), *Introduction to Logical Theory*, Methuen, London.

Veltman, Frank (1986), 'Data semantics and the pragmatics of indicative conditionals', in Elizabeth Traugott, Alice ter Meulen, Judy Reilly, & Charles Ferguson (eds), *On Conditionals*, Cambridge University Press, Cambridge, 147–68.

Warmbrod, Ken (1981a), 'Counterfactuals and substitution of equivalent antecedents', *Journal of Philosophical Logic*, 10, 267–89.

Warmbrod, Ken (1981b), 'An indexical theory of conditionals', *Dialogue, Canadian Philosophical Review*, 20, 4, 644–64.

Warmbrod, Ken (1982), 'A defense of the limit assumption', *Philosophical Studies*, 42, 53–66.

Warmbrod, Ken (1983), 'Epistemic conditionals', *Pacific Philosophical Quarterly*, 249–65.

Williams, Edwin (1974), 'Rule ordering in syntax', Ph.D. dissertation, MIT.

van der Wouden, Ton (1994), 'Negative contexts', Ph.D. dissertation, University of Groningen.

Wright, Crispin (1983), 'Keeping track of Nozick', *Analysis*, 43, 3, 134–40.

Zwarts, Frans (1995), 'Nonveridical contexts', *Linguistic Analysis*, 25, 286–312.

Zwarts, Frans (1997), 'Three types of polarity', in Fritz Hamm & Erhard Hinrichs (eds), *Plurality and Quantification*, Kluwer, Dordrecht.