Counterfactuals in a Dynamic Context

Kai von Fintel

1. INTRODUCTION

A primary goal of research in the semantics/pragmatics interface is to investigate the division of labor between the truth-conditional component of the meaning of an expression and other factors of a more pragmatic nature. One favorite strategy, associated foremost with Grice (1967, 1989), is to keep to a rather austere semantics and to derive the overall meaning of an utterance by predictable additional inferences, called "implicatures", which are seen as based on certain principles of rational and purposeful interaction. In this paper, I will explore a different way in which the truth-conditional component is complemented in context.

Imagine that we have persuasive evidence that an expression α in context c expresses a proposition P.

The straightforward way of capturing this in a semantic system is to attribute to α a context-dependent meaning that maps c to P in a systematic and adequate way.

(1)   <u>Analysis A</u>

$$[\![\alpha]\!]^c = P$$

A logically possible alternative attributes to α a meaning that has two aspects: (i) it alters the initial context c to a new context c′ and (ii) maps c′ to the proposition P in a systematic and, importantly, simpler way than under the first analysis.[1]

(2)   <u>Analysis B</u>

(i) $c \mid \alpha \mid = c$

(ii) $[\![\alpha]\!]^c = P$

I will argue for Analysis B in the case of counterfactual conditionals. But this is not easy. It should be clear that this kind of proposal cannot claim any conceptual advantages.[2] Even if Analysis B succeeds in making Step (ii), the truth-conditional component of the meaning of α, admirably

---

[1] I will adopt the perhaps risky strategy of supplementing the discussion in the text with sketches of formal machinery without commenting much on the technicalities. I hope that the formally adept readers won't need much handholding, while others will be able to grasp what is going on without studying the formalism. I will use the familiar notation $[\![\alpha]\!]^c$ for the denotation of α with respect to the contextual parameter c. The context-change potential of α is written as $\mid\alpha\mid$ and is a function from contexts to contexts. I follow common practice in dynamic semantics by writing the function to the right of its argument to symbolize the sequential nature of discourse.

[2] The Gricean strategy mentioned in the beginning is sometimes claimed to have a conceptual advantage (Grice speaks of a "modified Occam's Razor" that prefers an austere semantics). But I fail to understand why a two factor analysis should be methodologically preferred over a one factor analysis. The substantial support for Gricean analyses comes from empirical arguments that show the different behavior of truth-conditional aspects and "pragmatic" aspects (implicatures etc.).

austere, it would still appear to be disfavored in matters of overall complexity. Furthermore, it involves a somewhat strange conception of context change. The usual kind of context change contemplated in dynamic systems for natural language interpretation concerns the effect successful assertions have on the "conversational record" (introduction of new discourse referents, elimination of alternatives under consideration, etc.). What is going on according to Analysis B is that even before we compute the denotation of an expression it is able to change the input context. We will see that there is nevertheless a perfectly reasonable perspective on the context change posited here: context repair in the face of presupposition failure. We will also see a way of implementing the proposal in a dynamic semantics.

Critical support for the analysis has to come from empirical arguments. Perhaps, we have direct evidence for the reality of the austere semantics, so that the context-changing part is a necessary evil. Perhaps, we can find evidence that the context-changes posited by the analysis are in fact going on.

2.   THE NON-MONOTONICITY OF COUNTERFACTUALS

We will be concerned with examples like these:

(3)   a.   If kangaroos had no tails, they would topple over.
         b.   If Hoover had been a Communist, he would have been a traitor.
         c.   If Oswald hadn't killed Kennedy, someone else would have.

A pleasingly simple analysis of such counterfactuals with the schema *if ϕ, (then) would ψ*, from now on symbolized as $\phi > \psi$, would say that they are true at a world w iff all ϕ-worlds accessible from w are ψ-worlds.

(4) $[\![\phi > \psi]\!]^{f} (w) = 1$ iff $\forall w \in f(w): [\![\phi]\!]^{f} (w) = 1 \rightarrow [\![\psi]\!]^{f} (w) = 1$

         f: an accessibility function from worlds to sets of worlds.

This strict conditional analysis of counterfactuals runs into all kinds of trouble (for a classic discussion of some of the issues, see Goodman 1947). In particular, it turns out to be impossible to maintain a stable view of what accessibility function is needed in a given context. The problem can be illustrated by looking at the inference pattern of Strengthening the Antecedent.

(5)   *Failure of Strengthening the Antecedent (Downward Monotonicity)*

       If kangaroos had no tails, they would topple over.
       ⇏ If kangaroos had no tails but used crutches, they would topple over.

If one is in a somewhat silly frame of mind, one would presumably accept as true David Lewis' classic example *If kangaroos had no tails, they would topple over*. According to the strict conditional analysis this sentence claims that all of the worlds accessible from our world where kangaroos have no tails are worlds where they topple over. Imagine that a useful notion of accessibility here is the one that relates to our world all the worlds that share with it certain facts of a biological nature (and some laws of physics such as gravity) – but only certain ones: there must be worlds where kangaroos can live without tails. Among those "biologically accessible" worlds, all the ones where kangaroos have no tails are worlds where they chronically topple over. Since the claim is made about <u>all</u> such worlds, it is implicitly made also about those among the accessible worlds where kangaroos use crutches. Thus it should *a fortiori* be accepted as true that *If kangaroos had no tails but used crutches, they would topple over*. This however is an eminently debatable claim, even if one accepts the initial counterfactual. The strict conditional analysis incorrectly predicts the validity of Strengthening the Antecedent.

By now, it is well-known what the solution to this puzzle has to be (Stalnaker 1968, 1975, 1981, 1984; Stalnaker and Thomason 1970; Lewis 1973a,b, 1981).[3] The strict conditional analysis is wrong about the truth-conditions expressed by a counterfactual conditional. Instead, the semantics of the counterfactual must be stated so as to select the worlds quantified over in a way that is sensitive to the antecedent proposition. The first counterfactual makes a claim only about those accessible worlds which are as similar to our world as is possible while making the antecedent true. Arguably, in such worlds kangaroos would not use crutches, hence they topple over. The antecedent of the second counterfactual on the other hand forces us to look at worlds where tailless kangaroos use crutches. We again will focus in on those worlds among these oddities which are as similar <u>as possible</u> to our beloved actual world. The two conditionals thus make claims about two entirely different sets of worlds, which means it is not surprising that Strengthening the Antecedent is not an inference pattern to put any trust in.

---

[3] For two useful overviews on conditional semantics, see Nute (1984) and Edgington (1995).

One well-known variant of the this by now standard possible worlds semantics for counterfactuals states the truth-conditions relative to a relation of comparative similarity between worlds. The central notion is $w' \leq_w w''$, which says that w' is more similar to w than w'' is to w. A counterfactual quantifies over a selected subset of the antecedent ϕ-worlds, namely those ϕ-worlds that are at least as similar to the evaluation world w as any other ϕ-world.[4] In addition to similarity, we also speak interchangeably of "closeness" or "nearness" to the evaluation world.

(6)   For any proposition p, any similarity relation $\leq$, and any world w:

$$\max_{\leq, w}(p) = \left\{ w : p(w) = 1 \ \& \ \forall w'' : p(w'') = 1 \rightarrow w \leq_w w'' \right\}$$

(7) $[\![\phi > \psi]\!]^{\leq}(w) = 1$ iff $\forall w : w \in \max_{\leq, w}\left([\![\phi]\!]^{\leq}\right) \rightarrow [\![\psi]\!]^{\leq}(w) = 1$

Counterfactual conditionals attempt to maximize similarity with the actual world to the degree that that is possible in view of the counterfactual antecedent, so the worlds quantified over are those antecedent worlds that are maximally similar to the actual world. Since similarity is obviously a context-dependent and vague notion, counterfactuals will inherit these characteristics.

A crucial property of this semantics is that the conditionals are correctly predicted to be <u>non-monotonic</u>. In addition to the failure of Strengthening the Antecedent, here are two more failures of monotone inferences:

---

[4] Formulating the semantics of counterfactuals in these terms is only possible under what Lewis calls the Limit Assumption (that there will always be such a set of closest antecedent worlds), which Lewis in fact rejects. Stalnaker, on the one 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. I side with Stalnaker, not the least because it makes life easier. For discussion, see Lewis (1973a) and Stalnaker (1984: Chapter 7, esp. pp. 140-142). Further arguments <u>against</u> the Limit Assumption can be found in Herzberger (1979) and Pollock (1976: pp. 18-20). Further arguments <u>for</u> the Limit Assumption can be found in Warmbrod (1982).

(8)   *Failure of the Hypothetical Syllogism (Transitivity)*

If Hoover had been a Communist, he would have been a traitor.
If Hoover had been born in Russia, he would have been a Communist.
⇏ If Hoover had been born in Russia, he would have been a traitor.

(9)   *Failure of Contraposition*

(Even) if Goethe hadn't died in 1832, he would still be dead now.
⇏ If Goethe were alive now, he would have died in 1832.[5]

Hypothetical Syllogism (Transitivity) fails because even if all the closest ϕ-worlds are ψ-worlds and all the closest ψ-worlds are χ-worlds, we are not necessarily speaking about the same ψ-worlds (the ψ-worlds that ϕ takes us to may be rather remote ones). So in the Hoover-example, we get the following picture: The closest ϕ-worlds in which Hoover was born in Russia (but where he retains his level of civic involvement), are all ψ-worlds in which he becomes a Communist. On the other hand, the closest ψ-worlds in which he is a Communist (but retaining his having been born in the United States and being a high level administrator) are all χ-worlds in which he is a traitor. The closest ϕ-worlds do not include the closest ψ-worlds, so the Transitive inference does not go through.

Contraposition fails because the assumption that the closest ϕ-worlds are ψ-worlds does not preclude a situation where the closest *non-ψ*-worlds are also ϕ-worlds. The selected ϕ-worlds in which Goethe didn't die in 1832 are all ψ-worlds where he dies nevertheless (well) before the present. But of course, the closest (in fact, <u>all</u>) *non-ψ*-worlds (where he is alive today) are also ϕ-worlds where he didn't die in 1832.

3.   THE ALTERNATIVE

We will take it for granted from now on that the standard non-monotonic semantics for counterfactuals correctly describes the truth-conditions of a counterfactual uttered in an initial context. The counterfactual is true at a world w iff all of the ϕ-worlds closest to w are ψ-worlds.

What we will explore is the possibility of an analysis of the B type sketched in the introduction. Do we arrive at these truth-conditions in a way that is more complex than assumed by the standard theory? The idea

---

[5] Due to Kratzer.

would be that a counterfactual $\phi > \psi$ has two aspects to its meaning: (i) it shifts the input context to one where the accessibility function f, one of the contextual parameters, has been altered so as to assign to any world at least some ϕ-worlds, namely the closest ones, (ii) with respect to the newly altered context, the counterfactual will then be evaluated as true at a world w iff all of the ϕ-worlds accessible via f from w are ψ-worlds.

This alternative account treats the selection of the domain of quantification as the outcome of a process separate from the quantificational claim expressed by the counterfactual modal operator. I claim that counterfactuals both change and make use of a contextual parameter which I will call the "modal horizon". This parameter is simply the accessibility function appealed to by the simple strict conditional analysis, but it is now subject to evolution over the course of a discourse. The modal horizon gradually widens as more and more (counterfactual) possibilities are considered. The evolution of the modal horizon is governed by the same kind of similarity measure that is employed by the standard account. Counterfactuals assert that all ϕ-worlds assigned to the evaluation world by the current modal horizon are ψ-worlds. With respect to an initial, null context, my counterfactuals will have exactly the same truth-conditions as those of the standard analysis. The crucial difference between the analyses surfaces only when we consider sequences of counterfactuals. In my account, the modal horizon evolves and is kept track of during such a sequence. In the standard account, counterfactuals are considered in splendid isolation from surrounding discourse.

As I hinted in the introduction, this kind of account needs strong supporting evidence. Two properties of my account turn out to be of particular empirical value. (i) There are essentially dynamic facts concerning the way the order of counterfactuals in a sequence matters to the coherence of the sequence and to the plausibility of arguments. These facts demonstrate the need for context-change in the semantics of counterfactuals. (ii) The modal horizon-based analysis allows a restricted notion of entailment (which I call "Strawson-entailment") which involves the assumption of a constant horizon. According to this notion, patterns like Strengthening the Antecedent, Hypothetical Syllogism, Contraposition (for which there are spectacular counter-examples) turn

out to be valid inference patterns. I will suggest that there is empirical evidence for the limited kind of validity of these patterns.[6]

Again, on purely conceptual grounds, this kind of analysis has no particular attraction. It may be hard to appreciate why such a project would be of any interest. The suspicion may be that either these are mere notational variants or that the analysis by moving to an ill-understood pragmatic component will lose some of the rigor and explicitness of the standard approach. Before I move on to the empirical evidence, let me quote two masters who are somewhat skeptical of "pragmatic" solutions. But keep in mind that they may not have in mind the kind of analysis explored here.

Lewis wrote (1973: 13):

> It is still open to say that counterfactuals are vague strict conditionals based on similarity, and that the vagueness is resolved - the strictness is fixed - by a very local context: the antecedent itself. That is not altogether wrong, but it is defeatist. It consigns to the wastebasket of contextually resolved vagueness something much more amenable to systematic analysis than most of the rest of the mess in that wastebasket.

---

[6] My analysis belongs to the ever growing body of work concerned with the interaction of semantics with facts about context-dependence and mechanisms of context change. The importance of these aspects of linguistic meaning was urged on us by the same authors that pioneered the possible worlds semantics for counterfactuals: Stalnaker (1972, 1973, 1974, 1975, 1978, 1988, 1996) and Lewis (1979a,b). Semantic theories that give prominence to the context-changing potential of linguistic expressions have since then been developed by Heim (1982, 1983, 1992), Veltman (1981, 1985, 1986, 1996), Landman (1986), Barwise (1987), Groenendijk & Stokhof (1991), and many other researchers.

There is a (not very close) affinity between the dynamic analysis of counterfactuals explored here and pragmatic defenses of material implication analyses of indicative conditionals, such as the one presented by Grice (1967, 1989) and refined especially in Jackson's work (1979, 1984, 1987, 1990). Other pragmatically informed analyses of indicative conditionals include Veltman (1986) and McCawley (1993: 548ff). There is also a (somewhat closer) affinity with pragmatic defenses of strict implication analyses of subjunctive conditionals or variants thereof, such as the ones presented by Warmbrod (1981a,b, 1983), Wright (1983), and Lowe (1990, 1995). Warmbrod's proposal in particular will be discussed as a precedent for my account.

Stalnaker also saw a "pragmatic" option (in fact, his analysis of indicative conditionals in his 1975 paper is a precedent for my analysis of counterfactuals) but was more open to this possibility than Lewis in the passage above. Stalnaker wrote (1987: 125f):

> One can defend a strict conditional account of conditionals against the counterexamples and arguments we have given by emphasizing the context-dependence of conditionals. One may argue that the conditional is *semantically* a fixed strict conditional but that the domain of possible worlds relative to which it is defined varies with context. Apparent failures of hypothetical syllogism and contraposition are to be explained as fallacies of equivocation caused by shifts in context. For example, one may say of a counterexample to hypothetical syllogism that the first premise seems true because it suggests one context, while the second seems true because it suggests a different context. Perhaps no single plausible context will be one relative to which both premises are true. Therefore, it is argued, the counterexamples do not defeat the claim that the inference is, within any single context, valid. … A suitable elaboration of this reply would build into the *pragmatics* of conditionals an apparatus similar to what is built into the *semantics* in the kind of theory I am defending. … Despite the differences in logic, the difference between a strict conditional theory and a theory of the general kind I am defending might be more superficial than it seems. The principal difference might be in where the line between semantics and pragmatics is drawn, which will determine at what level of abstraction one's notion of validity is defined. But the question is not arbitrary. If one draws the line in the wrong place, one may not only give a less efficient and perspicuous description of the phenomena, one may miss some significant generalizations. In general, if contexts shift too easily and often, then semantic validity will have little to do with the persuasiveness of arguments. Generalizations about the structure of arguments may be missed. On the other hand, if a simple semantics for a specific kind of construction, combined with general pragmatic principles governing the structure of discourse, can account for the complexities of the context shifts, one may have a better overall theory even if a purely semantic concept of validity loses its close connection with the phenomena of argument.

4. EVIDENCE

The alternative analysis will start earning its keep when we turn to the consideration of sequences of counterfactuals. A crucial feature of the account is that the modal horizon is passed on from one counterfactual to the next and that it continually evolves to include more and more possibilities. Where the analyses will crucially differ is in cases where a counterfactual early in a sequence has brought into play some remote possibility. According to the type B account defended here, a later counterfactual cannot ignore any possibilities as far out as the possibilities considered earlier. According to the standard account, the later counterfactual is allowed to just make a claim about the closest antecedent worlds and ignore any more remote possibilities.

One case in point comes from Lewis-Sobel sequences. Lewis argued against the view that in his counterexamples to Strengthening the Antecedent the context is subtly shifted. He in fact maintained that the pertinent examples are cases where the context (that is, the selection function or similarity measure) remains relevantly the same throughout the examples. He attempts to demonstrate this with the following kind of example (based on examples due to Sobel):

(10)   If the USA threw its weapons into the sea tomorrow, there would be war; but if all the nuclear powers 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 was meant to ensure that the context stays constant throughout, an assumption that in our more dynamic days seems rather naïve.

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.

Consider now the clear contrast between Lewis' example and a variant due to Irene Heim:

(11)   ??If all the 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 (11), 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.[7]

The fact that (10) 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 (10) the set of worlds quantified over properly expands, but that over the course of

---

[7] Braine (1979) has a different slant on the example. He thinks that even Lewis' original example is not acceptable. According to him, it is merely a very sloppy way of expressing the following perfectly acceptable statement:

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

Note that here a reversal of the order has no effect on the plausibility of the statement:

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

I don't agree with Braine's assessment of the original example: it is not a <u>sloppy</u> way of stating the first part of the assertion, it is merely a more <u>context-dependent</u> way of doing so. The reversal of the conjuncts changes the context and thus affects the meaning of the second conjunct, thus changing the meaning of the whole statement.

(11) it cannot shrink. This asymmetry is unexpected if one maintains there is no context change.

Note also that if someone utters (10), someone else can then rejoin that the initial conditional is "no longer" true:[8]

(12)   A: If the USA threw its weapons into the sea tomorrow, there would be war; but if all the nuclear powers 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.[9]

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

This is unexpected under a 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 the nuclear powers become meek. But that doesn't seem to happen.[10]

---

[8] 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 may cease to be true in a later context where the standards of precision have been sharpened.

[9] 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 a separate paper of mine (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.

[10] Note that B' in her reply to A 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 will be valid in mine (if we bothered to introduce *might*-counterfactuals explicitly).

The same kind of sensitivity to the order in which conditionals are presented that we observed in (10) vs. (11) can be detected when we examine the Hoover counterexample to HS. Think about the example again:

(8)   If Hoover had been a Communist, he would have been a traitor.
         If Hoover had been born in Russia, he would have been a Communist.
         ⇏ If Hoover had been born in Russia, he would have been a traitor.

Now note what happens when we reverse the order in which the premises are presented:

(13)   If Hoover had been born in Russia, he would have been a Communist.
         ??If Hoover had been a Communist, he would have been a traitor.
         ⇏ If Hoover had been born in Russia, he would have been a traitor.

We are not at all tempted to admit both premises as true. The natural way of reading the second premise is as taking into account a set of Communist scenarios <u>including</u> those introduced by the first premise. With respect to the active context then, the second premise is naturally read as expressing a proposition that is in fact false.

We have now seen some reasons that the continual context changes posited by my account are in fact observable. The other kind of motivational evidence for the analysis comes from arguments that the austere statement of the truth-conditions is supported by facts about grammar. Consider negative polarity items (NPIs), expressions that are prototypically allowed in the semantic scope of negation but not in "positive" environments. Two examples are *any* and *ever*:

(14)   a. I don't think we have any potatoes.
             *I think we have any potatoes.
         b. I don't think there will ever be another Aristotle.
             *I think there will ever be another Aristotle.

Ladusaw (1979) showed that NPIs are licensed in downward monotone positions. The most spectacular illustration can be found in quantified sentences:

(15)   a. *Some (student who has ever been to Rome) (has liked it there).
         b. No (student who has ever been to Rome) (has liked it there).
         c. Every (student who has ever been to Rome) (has liked it there).

The determiner *some*, which is not downward monotone in either argument, does not license NPIs. The determiner *no*, which is downward

monotone in both its arguments, licenses NPIs in both positions. The determiner *every* is downward monotone in its first argument and licenses NPIs there, but is upward monotone in its second argument and does not license NPIs there. Now, NPIs <u>are</u> licensed in the antecedent of conditionals:[11]

(16)   If you had left any later, you would have missed the plane.

This fact has always been problematic from the point of view of Ladusaw's generalization, since it was also accepted that conditional antecedents are not downward monotone contexts (Strengthening the Antecedent is after all invalid). Kadmon & Landman (1993) suggest that conditionals are in fact downward monotone, as long as we keep the context constant for the whole stretch of the argument. The same idea is advocated by Katz (1991). These authors fail to address the fact that the Stalnaker-Lewis analysis (and Kratzer's variant, which those authors primarily refer to) actually does claim to keep the value of contextually supplied parameters constant. The similarity measure is the same for both premise and conclusion. Strengthening the Antecedent is invalid because the measure can react very differently to the two antecedents (the original one and the strengthened one). It would be nice if we had a semantics of conditionals that gave us some kind of limited monotonicity to plug into the general theory of NPI-licensing. I suggest that the analysis I'm exploring here is up to this task: we will be able to formulate a limited kind of entailment, with respect to which counterfactual antecedents will be downward monotone environments hospitable to NPIs.[12]

This concludes the initial review of the kind of evidence that supports an analysis that combines automatic context-change with a strict conditional statement of the truth-conditions of counterfactuals. We turn now to the formal development of the analysis.

---

[11] Partee (1993) 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.

[12] Heim (1984) presents an attempt at defining a limited kind of downward monotonicity to license NPIs, but her proposal is shown to be insufficient by Kadmon & Landman (1993).

5. WARMBROD

Our starting point comes from the work of Warmbrod (1981a,b), who formulates his proposal in terms of a strict conditional analysis which interprets counterfactuals relative to a contextually given accessibility relation R. "The particular relation R that we use in deciding on the truth-value of a given conditional may be thought of as varying from one occasion of speech to another. So it is immediately apparent that a large task remains for pragmatic theory. Listeners usually seem to know how to interpret the conditionals of everyday discourse, and hence we must assume that they know what accessibility relation is to be used in interpreting conditionals. Hence, we need a pragmatic theory of interpretation to explain how they manage to identify the right relation." Determining the accessibility relation is a task that needs to be done once and for all for all the conditionals in a piece of coherent discourse. A conditional is always evaluated relative to a whole piece of discourse rather than in isolation. "When several conditionals appear together in a single corpus, it seems … reasonable that we identify a single relation … early in the corpus. That relation is then held constant until the corpus ends" (1981a: 279). Based on an early (I guess, ideally, the first) conditional in a discourse, a "standard" accessibility relation is chosen as follows:

(17)   An accessibility relation R results from a *standard interpretation* of an antecedent ϕ in w relative to a comparative similarity relation $\leq$ iff
wRw iff $\forall w'' \in [\![\phi]\!]$: $w \leq_w w''$.

The accessibility relation should only get us to worlds that are at least as similar to w as the most similar antecedent worlds. Warmbrod's "formalism captures the same basic intuition (for isolated counterfactuals) that Stalnaker and Lewis have in mind" (1981a: 280). For discourses containing more than one counterfactual, Warmbrod imposes a condition of "normality":

(18)   An accessibility relation R is *normal* for a body of discourse D relative to a world w iff for every antecedent ϕ in D, there is some ϕ-world w' such that wRw'.

Warmbrod's "pragmatic theory of interpretation embodies two claims. First, a relation R used to interpret a body of discourse D must result from standard interpretation of some antecedent or hypothesis advanced early in D. Second, R must be normal for D relative to the actual world" (1981a: 282).

This analysis has some of the ingredients of the approach we are planning to develop. It combines a strict conditional analysis with a view of the context that admits that the context may change. Warmbrod does not give a compositional procedure for context-change. What he does is state a global condition under which a context does <u>not</u> change in a sequence of counterfactuals. He requires the context to stay constant for a well-behaved, "normal" discourse. The global nature of his normality condition prevents him from dealing with embedded conditionals (as he readily admits in a footnote). While assigning an important role to context-change in the analysis, Warmbrod puts the emphasis on defending the strict conditional semantics.

6. HEIM

Some of the technical deficiencies of Warmbrod's proposal are not found in a system sketched by Irene Heim in a presentation in an MIT seminar in the spring of 1994 (her goal was to put on firmer ground the idea found in Katz (1991) that conditionals are downward monotonic <u>if</u> one keeps the context constant). Heim adopts a strict conditional semantics as far as truth-conditions are concerned but adds to this a semantic presupposition about the contextually supplied accessibility function: f has to assign to the evaluation world a set of worlds containing at least some antecedent worlds.[13]

(19)   <u>Heim's Semantics for Counterfactuals</u>

f is an accessibility function and $\leq$ a comparative similarity order.

a.   Compatibility Presupposition

$[\![\phi > \psi]\!]^{f, \leq}$ is defined at w only if $\exists w \in f(w): [\![\phi]\!]^{f, \leq}(w) = 1$

b.   Truth-Conditions

If defined at w,
$[\![\phi > \psi]\!]^{f, \leq}(w) = 1$ iff $\forall w \in f(w): [\![\phi]\!]^{f, \leq}(w) = 1 \rightarrow [\![\psi]\!]^{f, \leq}(w) = 1$

---

[13] This analysis turns Warmbrod's global normality requirement into a semantic presupposition of the counterfactual and will therefore be able to deal with embedded conditionals.

Two important issues are left open by this semantics: (i) how is the accessibility function initially identified? (ii) and what happens when an initially identified accessibility function does not include accessible antecedent worlds for some counterfactual later on? These issues need to be addressed in a separate part of the theory. Heim herself, in her class handout, merely stated that the accessibility function needs to deliver a similarity-based Lewis-sphere around the evaluation world.

(20)   Admissible Contexts (Lewis-Sphere Condition)

A context with the parameters f and ≤ is only admissible if
$$\forall w: \forall w \in f(w): \forall w'': w'' \leq_w w \rightarrow w'' \in f(w)$$

I would like to propose that we add two further ingredients to the theory of contextually supplied accessibility functions: (i) Initially, we assume a trivial accessibility function which assigns to any evaluation world w world merely the singleton set of that world itself *{w}*. (ii) Any time the compatibility presupposition fails (which will be the case for any initial counterfactual under the previous assumption), the context is adjusted by minimally expanding the accessibility function so as to satisfy the compatibility function. The role of the similarity measure will crucially lie in controlling the initial identification of the accessibility function and its subsequent evolution. Instead of formalizing this here, I will do so in the next section in a more fully dynamic system.

Before we take on that task, let me assess whether this semantics is monotonic or non-monotonic. In fact, adding a compatibility/normality presupposition to a strict conditional semantics makes it just as non-monotonic as the standard account. For example, Strengthening the Antecedent is invalid: for a given world w the premise $\phi > \psi$ may be true (with respect to the contextual parameter f) while the conclusion $\phi \, \& \, \chi > \psi$ may be undefined (because f might assign to w a set of worlds none of which is a $\phi \, \& \, \chi$-world).

But the new semantics allows us to formulate a special notion of entailment according to which counterfactuals are monotonic. Heim, who wanted to make counterfactuals monotonic enough to license NPIs, suggested that if one requires a constant context and requires all the propositions in the argument to be defined (all presuppositions have to be satisfied), one validates e.g. Strengthening the Antecedent in a limited way. This is parallel to Warmbrod's requirement for the context to be "normal" for a whole sequence.

(21)   <u>Strawson-Entailment à la Heim</u>

$\phi_1, \ldots, \phi_n \vDash_{\text{Strawson}} \psi$ iff for all f, ≤, w such that

$[\![\phi_1]\!]^{f, \leq}, \ldots, [\![\phi_n]\!]^{f, \leq}$, and $[\![\psi]\!]^{f, \leq}$ are all defined at w

if $[\![\phi_1]\!]^{f, \leq}(w) = 1, \ldots, [\![\phi_n]\!]^{f, \leq}(w) = 1$, then $[\![\psi]\!]^{f, \leq}(w) = 1$

The definedness condition here is what enforces the accessibility function to be lax enough to accommodate all the antecedents considered in the sequence of counterfactuals, in particular the conclusion.

Here is why I have called this notion "Strawson-entailment". Strawson (1952) in a famous passage discusses the possibility of making the traditional "subaltern" 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 what is now called a semantic or logical presupposition: if the presupposition is not satisfied, the sentence will be neither true nor false. The subaltern inference is straightforwardly valid in this system: whenever *Every S is P* is true its existence presupposition, that there are Ss, must be satisfied, but then *Some S is P* will have to be true as well. Unfortunately, some other traditional inferences will now be valid. For example, the traditional simple conversion of negative universal statements will fail. *No S is P* may be true (there are Ss but none of them are Ps) while *No P is S* may be neither true nor false (there are no Ps). Strawson then suggests that when in traditional logic the inference above is called valid what is 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: p. 176). There you go. This will indeed solve Strawson's problem.

If Strawson was right, this notion of entailment is natural enough to have engendered an entire tradition of logic (the one that assumed existential import for natural language quantifiers). And if Heim is right, this notion of entailment underlies grammatical NPI-licensing. Of course, one may well be skeptical of Strawson's own reason for adopting this notion while following Heim in taking it to be operative in cases where it enforces

constancy of a particular contextual parameter. The latter is surely a feature often assumed to be essential for proper logical argumentation.[14]

We will now explore the possibility of implementing what for Heim is a task for pragmatics (the selection of a new context once the current one has led to presupposition failure) into the "context change potential" of counterfactuals.

7. DYNAMIC IMPLEMENTATION

The story in Heim's sketch is that counterfactuals presuppose that the contextually supplied accessibility function contains worlds at which the antecedent is true. Whenever a previously uncontemplated antecedent possibility is introduced, there is presupposition failure and a decision has to be made whether the context should be adjusted by adding the novel antecedent to the set of relevant possibilities. If the decision is made to adjust the context, this is most economically done by "expanding" the accessibility function just enough to encompass some worlds at which the antecedent is true.

What is done in a dynamic semantic system is to encode the effect a successful assertion of a sentence has on the context in the "dynamic meaning" of the sentence. Well-known dynamic effects concern the progressive elimination of worlds from the context set (the set of worlds which according to the common ground of the conversation are still viable candidates to be the actual world) and the addition of discourse referents to the domain of entities that can be referred to anaphorically. What we are concerned with here is the gradual "expansion" of the accessibility function, which is a contextual parameter of evaluation for counterfactual sentences. As mentioned earlier, to have a more evocative name for this evolving contextual parameter, I will call it the "modal horizon".

The procedure will be this: If a conditional is accepted as an assertion, the context will first be changed to expand the modal horizon if the antecedent wasn't already considered a relevant possibility. Then, the conditional will be interpreted in the new context. What we would like to do then is to assign to the counterfactual $\phi > \psi$ a context change potential, a function from contexts to contexts that changes the context so as to add

---

14 Elsewhere, I explore Strawson-Entailment and NPI-Licensing in detail (von Fintel 1999).

the antecedent to the modal horizon. The proposition expressed by the conditional is then computed with respect to the already updated context. Somewhat formally:

(22)   Dynamic Semantics for Counterfactuals (Rough)

   a.   Context Change Potential

$$f \mid \phi > \psi \mid^{\leq} = \lambda w.\, f(w) \cup \left\{ w : \forall w'' \in [\![\phi]\!]^{f,\,\leq} : w \leq_w w'' \right\}$$

   b.   Truth-Conditions

$$[\![\phi > \psi]\!]^{f,\,\leq}(w) = 1 \text{ iff}$$
$$\forall w \in f \mid \phi > \psi \mid^{\leq}(w) : [\![\phi]\!]^{f,\,\leq}(w) = 1 \rightarrow [\![\psi]\!]^{f \mid \phi > \psi \mid^{\leq},\,\leq}(w) = 1$$

We pretend that the modal horizon f is the only contextual parameter which evolves in the course of a sequence. Sentences are assigned as context change potentials functions from input modal horizons to new potentially updated modal horizons. The only sentences which effect any context change are counterfactuals (we pretend that apart from atomic sentences, negated sentences, and conjunctions, we have only counterfactuals). These update f by adding to it for any world w the closest antecedent worlds. Of course, if such worlds are already assigned by f to w, the new f will be the same as the old f. To assess the truth of a counterfactual at a world w with respect to an initial f and a similarity relation $\leq$, we first update f with the context-change potential of the counterfactual. Then, we check whether all of the worlds assigned to w by the updated f which are ϕ-worlds are also ψ-worlds.

This rough first attempt has no provision for embedded conditionals. If the antecedent contains a conditional, we should presumably allow the embedded conditional to update the accessibility function before we look for the worlds quantified over by the containing conditional. If the consequent contains a conditional, we will have to pass on to subsequent discourse the update that this embedded conditional effects. We revise and clean up the proposal:

(23)   <u>Dynamic Semantics for Counterfactuals (Revised)</u>

a.   Auxiliary Notion: Update of f by a sentence ϕ

For any sentence ϕ, any accessibility function f, and similarity relation $\leq$:
$$f^{\phi, \leq} = \lambda w.\, f(w) \cup \left\{ w : \forall w'' \in [\![\phi]\!]^{f, \leq} : w \leq_w w'' \right\}$$

b.   Context Change Potential of Counterfactuals

$$f \mid \phi > \psi \mid^{\leq} = f^{\phi, \leq} \mid \phi \mid^{\leq} \mid \psi \mid^{\leq}$$

c.   Truth-Conditions

$$[\![\phi > \psi]\!]^{f, \leq}(w) = 1 \text{ iff}$$
$$\forall w \in f^{\phi, \leq} \mid \phi \mid^{\leq}(w): [\![\phi]\!]^{f, \leq}(w) = 1 \rightarrow [\![\psi]\!]^{f^{\phi, \leq} \mid \phi \mid^{\leq}, \leq}(w) = 1$$

Over the course of a piece of reasoning/discourse in which a number of conditionals are asserted the context will naturally evolve so as to expand the modal horizon. If the semantics in (23) is all there is, we have a one-way street: more and more possibilities are introduced. One can think of the modal horizon f as a discourse referent but one that continues to be updated throughout a discourse. It is like a mailbox into which more and more items are stuffed. It is in this respect that the present account differs from the standard account in which the set of worlds selected by the similarity measure shrinks and expands according to the whim of the antecedent.

The relocation of non-monotonicity from the semantics of the modal operator to the contextual evolution of the modal horizon does not lead to a difference in the truth-conditions derived for an isolated counterfactual. Assume that initially, f is trivial in that it assigns to each world w only $\{w\}$. Now, $\phi > \psi$ is offered. f needs to be expanded. Apart from w, we need to have in $f(w)$ the closest ϕ-worlds and all additional non-ϕ-worlds that are closer to w than the closest ϕ-worlds. The conditional now claims that all of the ϕ-worlds in $f(w)$ are ψ-worlds. The ϕ-worlds in $f(w)$ at this point are exactly the closest ϕ-worlds to w.

It is once we move beyond isolated conditionals uttered in a null context, the approaches will begin to diverge. The central point of this paper is that we now have an explanation both of the Stalnaker-Lewis counterexamples

to the classic inference patterns <u>and</u> of the data discussed earlier. Before we turn to judgments and intuitions about coherence and entailment, we will have to question the one-way nature of the evolution of the modal horizon.

8. ONE-WAY STREET?

One-way expansion in modalized arguments is discussed in Lewis' seminal paper on score-keeping (Lewis 1979b). He writes:

> Suppose I am talking with some elected official about the ways he might deal with some embarrassment. So far, we have been ignoring those possibilities that would be political suicide for him. He says: 'You see, I must either destroy the evidence or else claim that I did it to stop Communism. What else can I do?' I rudely reply: 'There is one other possibility - you can put the public interest first for once!' That would be false if the boundary between relevant and ignored possibilities remained stationary. But it is not false in its context, for hitherto ignored possibilities come into consideration and make it true. And the boundary, once shifted outward, stays shifted. If he protests 'I can't do that', he is mistaken.

Lewis there also makes a similar observation about rapid context shifts in the analysis of statements about knowledge, which then became the topic of a more recent paper (Lewis 1996). Note that Lewis clearly articulates the one-way nature of the context shift (more possibilities come into consideration, no shrinking back allowed).[15]

Is the expansion of the modal horizon really irreversible? It seems that there should be procedures for "resetting" of the context. Imagine a speaker deliberating as follows:

---

[15] Of course, we have to note that the expansion in this example is not one we can account for here. Lewis' example involves a modal statement "You can put the public interest first for once", which goes beyond our simple system. One obvious way of incorporating such sentences would be to have *can* ϕ claim that there are ϕ-worlds in the current modal horizon. If the current horizon does not in fact support this claim, one way of repairing the situation is by expanding the horizon. Further development of this idea will have to await a future occasion.

(24)   If the USA threw its weapons into the sea tomorrow, there would be war. Well, if all the nuclear powers threw their weapons into the sea tomorrow, there would be peace. But of course, that would never happen. So, as things stand, if the USA threw its weapons into the sea tomorrow, there would be war.

It appears that this resetting of the context has to rely on explicit indications, while expansion occurs silently and smoothly. Consider the interpretation of the crucial resetting sentence:

(25)   But of course, that would never happen.

I take it that the most natural way of reading this sentence is that it asserts that the actual world w is such that no worlds in $f(w)$ are worlds where all the nuclear powers throw their weapons into the sea tomorrow. So, the claim is that this is not a relevant possibility. Now, by the time the second counterfactual in (24) has been processed and accepted, the modal horizon f will have been such that it assigns to any world a set of worlds some of which are in fact such worlds. So, the claim made by (25) would be blatantly false in that context. There are two obvious ways of incorporating the resetting effect of (25): (i) one could posit a reset operator that is prefixed to such sentences and provide it with a semantics that has it eliminate possibilities from the modal horizon, or (ii) one could treat resetting as a pragmatic operation that occurs at a higher level than the dynamic semantics, essentially as a repair mechanism. Perhaps, it is the fact that (25) is in blatant contradiction with the preceding discourse that will trigger a contraction of the modal horizon.

One piece of evidence that at least sometimes resetting is a rather indirect pragmatic mechanism comes from examples like this one:

(26)   A: If John had been at the party, it would have been much more fun.

        B: Well, if John had been at the party and had gotten into a fight with Perry, that wouldn't have been any fun at all.

        A: Yes, but Perry wasn't there. So, if John had been at the party, he wouldn't have gotten into a fight with Perry.

Here, the factual assertion that Perry wasn't there seems to trigger a resetting of the context (or serve as a reason for not admitting in the first place the expansion proffered by B). Of course, factual assertions cannot in general have such a disruptive effect on counterfactual reasoning, which precisely serves to abstract away from (certain) facts. The following is a case in point:

(27)   A: If John had been at the party, it would have been much more fun.

        B: But John wasn't at the party.

        A: Yes. I said if he <u>had</u> been there, it would have been more fun.

For the time being, I do not have a theory of when and how the modal horizon can be expanded and contracted by expressions other than conditionals. I need to leave this interesting topic for some other occasion.[16]

9. INFERENCES AND VALIDITY

In classic logic, it is considered imperative that in the assessment of arguments the context remain stable. But once we recognize the dynamic character of language - the fact that the context changes all the time and that language is one of the driving forces of this context change - this changeability of the context may need to be taken into account when we talk about the validity of inferences.[17]

A dynamic notion of entailment already appears in Stalnaker's paper "Indicative Conditionals" (1975), one of the founding papers of dynamic semantics. Stalnaker argues that the inference from *ϕ or ψ* to *if not ϕ, then ψ* (indicative), which is invalid in his semantics, is nevertheless a "reasonable inference" as he defines that notion. We can translate his notion into our system as follows:

(28)   <u>Dynamic Entailment</u>

$\phi_1, ..., \phi_n \vDash_{\text{dynamic}} \psi$ iff for all contexts c,

$[\![\phi_1]\!]^c \cap ... \cap [\![\phi_n]\!]^{c|\phi_1|\cdots|\phi_{n-1}|} \subseteq [\![\psi]\!]^{c|\phi_1|\cdots|\phi_n|}$

---

[16] Context resetting will presumably be more pervasive when we try to apply the methods of this paper to indicative/epistemic conditionals. I hope to turn to this task soon in a companion paper. Both Jeroen Groenendijk and Manfred Krifka have urged me to think more about cases of contraction of the modal horizon.

[17] There is some technical discussion of various possible notions of validity in dynamic systems, see especially van Benthem (1995), Dekker (1996), Muskens et.al. (1997), and Veltman (1996).

For any starting context it has to hold that when the premises are successively asserted and accepted, the context evolves into one whose context set is included in the proposition expressed by the conclusion in that resulting context. There is a corresponding notion of dynamic consistency:

(29)   <u>Dynamic Consistency</u>

A sequence $\phi_1, \ldots, \phi_n$ is dynamically consistent iff there is a context c s.t.
$$\llbracket \phi_1 \rrbracket^c \cap \ldots \cap \llbracket \phi_n \rrbracket^{c \left| \phi_1 \right| \cdots \left| \phi_{n-1} \right|} \neq \varnothing$$

Consider as a useful application of this dynamic notion of consistency the difference between the Lewis-Sobel sequence and Heim's variant:

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

(11)   ??If all the 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.

The Lewis-Sobel sequence is dynamically consistent because we can start with a context whose modal horizon is just wide enough to include those ϕ-worlds that are ψ-worlds; this horizon is then widened by the second sentence, which may well be true if all of the closest $\phi\&\chi$ -worlds that now come into vision are *non-ψ*-worlds. The Heim sequence is dynamically inconsistent, because we have no automatic mechanism that would allow the horizon to shrink between the addition of the first sentence and the assessment of the second sentence. And so, the first sentence makes the claim that the $\phi\&\chi$-worlds in the set of accessible worlds are all *non-ψ*-worlds, while the second sentence makes the claim that all the ϕ-worlds in the very same set of accessible worlds are ψ-worlds: a straightforward contradiction.

The notion of dynamic entailment is one which replicates in my system most of the logical assessments made by the standard theory. Most of the classic monotonic inference patterns are dynamically invalid. There is one notable exception: Hypothetical Syllogism is dynamically invalid when the premises are ordered in one way: $\psi > \chi, \phi > \psi \therefore \phi > \chi$; but it is dynamically valid when the premises come in the other order: $\phi > \psi, \psi > \chi \therefore \phi > \chi$. In effect the, exactly those patterns are dynamically invalid for which we can find intuitive counterexamples. While this seems nice, there

is reason to think that dynamic entailment is in fact not the notion that we use to assess logical arguments. Consider the following sequence as an example of dynamic entailment:

(30)   If the USA threw its weapons into the sea tomorrow, there would be war; but if all the nuclear powers threw their weapons into the sea tomorrow, there would be peace. So, if the USA threw its weapons into the sea tomorrow, there might not be war.

This argument uses as the sequence of premises the same Lewis-Sobel sequence that we saw to be dynamically consistent. By the time the second premise has widened the modal horizon, the conclusion will be true with respect to that newly expanded context. So, the argument in (30) is dynamically valid.

But of course, (30) is quite weird when seen not as coming from a speaker whose mind is evolving but as a deliberate argument for a particular conclusion.[18] This suggests to me that while the dynamic notion of consistency is quite the correct tool for assessing evolving discourse, for assessing <u>logical arguments</u> for a particular conclusion the fully dynamic notion of entailment is not adequate. For logical argumentation, we take speakers to be committed to a stable context. Someone who makes a logical argument gives an implicit promise that the context is not going to change during the argument. We should say that a speaker who presents a sequence as an attempt to argue for a particular conclusion has to be assuming a context which is such that the argument does not change the modal horizon.

This introduces in the context of our dynamic system a notion of entailment corresponding to Strawson-entailment as discussed earlier:

(31)   <u>Strawson-Entailment</u>

$\phi_1, \ldots, \phi_n \vDash_{Strawson} \psi$ iff for all contexts c such that $c = c \left| \phi_1 \right| \ldots \left| \phi_n \right| \left| \psi \right|$,

it holds that $[\![\phi_1]\!]^c \cap \ldots \cap [\![\phi_n]\!]^{c \left| \phi_1 \right| \cdots \left| \phi_{n-1} \right|} \subseteq [\![\psi]\!]^{c \left| \phi_1 \right| \cdots \left| \phi_n \right|}$

Consider as an example the validity of Strengthening the Antecedent (SA). Under the <u>assumption</u> that a context is already such that *if ϕ* and *if $\phi \& \chi$* will not further expand the accessibility function, the inference is fine:

---

[18] Thanks to Jeroen Groenendijk for this observation and the example in (30).

thus, it is Strawson-valid. It appears to me that speakers can reasonably offer arguments of the form of Strengthening the Antecedent. What should we say about such behavior? Do they commit a fallacy? More likely, what we want to say is that they must be making tacit additional assumptions that make their inference valid. According to my account, the additional assumption that they are making is that the accessibility function is such that it remains constant throughout the inference.

As we saw earlier, there is another reason to explore the usefulness of "Strawson-entailment". There is a well-established account according to which the validity of "downward" inferences such as Strengthening the Antecedent is the factor that governs the appearance of *negative polarity items* (NPIs) in various environments. The antecedent of conditionals is a well-known problem case for this account. If we could pursue an account according to which Strawson-entailment is the operative notion in NPI-licensing, we might get somewhere interesting (see von Fintel 1999).

10. OUTLOOK

This paper has presented a sketch of an alternative implementation of the standard possible worlds semantics of counterfactuals. I would like to end by briefly pointing out some respects in which the account sketched here should be further investigated. (i) I have kept away from exploring the internal compositional semantics of conditionals. Kratzer (1977, 1978, 1979, 1981a,b,c, 1986, 1991) has convincingly argued that we need to derive the meaning of conditionals from independently motivated analyses of modal operators (such as *would*) in combination with *if*-clauses. It remains to be seen how to best translate my sketch into her richer system. (ii) One important extension would be to apply the system to indicative/epistemic conditionals. This will be more complex than what I did here because there will have to be some tighter interaction between the evolution of the context set and that of the modal horizon; see von Fintel (1998) for some initial ideas about the indicative/subjunctive distinction, without context-dynamics however. (iii) One should explore the connections to other dynamic or DRT-based approaches to modality (especially Roberts 1986, 1989, 1995; Kasper 1992; Kibble 1994, 1996; Geurts 1995; Frank 1997; Stone 1997). (iv) My system does not incorporate provisions for presupposition projection. One should try to see how much of Heim's suggestions can be carried over (Heim 1992).

REFERENCES

Barwise, Jon (1987), 'Noun Phrases, Generalized Quantifiers, and Anaphora', in Peter Gärdenfors (ed.), *Generalized Quantifiers*, Reidel, Dordrecht, 1-29.

van Benthem, Johan (1995), *Language in Action: Categories, Lambdas and Dynamic Logic*, MIT Press, Cambridge.

Dekker, Paul (1996), 'The Values of Variables in Dynamic Semantics', *Linguistics and Philosophy*, **19**, 211-257.

Edgington, Dorothy (1995), 'On Conditionals', *Mind,* **104**, 235-329.

von Fintel, Kai (1997), 'Bare Plurals, Bare Conditionals, and *Only*', *Journal of Semantics,* **14**, 1-56.

von Fintel, Kai (1998), 'The Presupposition of Subjunctive Conditionals', in Uli Sauerland and Orin Percus (eds.), *The Interpretive Tract,* MITWPL, Cambridge, 29-44.

von Fintel, Kai (1999), 'NPI-Licensing, Strawson-Entailment, and Context-Dependency', *Journal of Semantics.*

Frank, Annette (1997), *Context Dependence in Modal Constructions*, Ph.D. Dissertation, Universität Stuttgart.

Geurts, Bart (1995), *Presupposing*, Ph.D. Dissertation, Universität Stuttgart.

Goodman, Nelson (1947), 'The Problem of Counterfactual Conditionals', *Journal of Philosophy,* **44**, 113-128.

Grice, Paul (1967), 'Logic and Conversation', ms.

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

Groenendijk, Jeroen and Martin Stokhof (1991), 'Dynamic Predicate Logic', *Linguistics and Philosophy,* **14**, 39-100.

Heim, Irene (1982), *The Semantics of Definite and Indefinite Noun Phrases*, Ph.D. Dissertation, University of Massachusetts at Amherst.

Heim, Irene (1983), 'On the Projection Problem for Presuppositions', *WCCFL,* **2**, 114-125.

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.

Jackson, Frank (1979), 'On Assertion and Indicative Conditionals', *Philosophical Review,* **88**, 565-589.

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**, 134-147.

Kadmon, Nirit and Fred Landman (1993), '*Any*', *Linguistics and Philosophy,* **16**, 353-422.

Kasper, Walter (1992), 'Presuppositions, Composition, and Simple Subjunctives', *Journal of Semantics,* **9**, 307-331.

Katz, Graham (1991), 'The Downward Entailingness of Conditionals and Adversatives', *FLSM,* **2**, 217-243.

Kibble, Rodger (1994), 'Dynamics of Epistemic Modality and Anaphora', in Harry Bunt, R. Muskens, and G. Rentier (eds.), *Proceedings of the International Workshop on Computational Semantics*, ITK, Tilburg.

Kibble, Rodger (1996), 'Modal Subordination, Focus, and Complement Anaphora', *Proceedings of the Tbilisi Symposium on Language, Logic, and Computation.*

Kratzer, Angelika (1977), 'What *Must* and *Can* Must and Can Mean', *Linguistics and Philosophy,* **1**, 337-355.

Kratzer, Angelika (1978), *Semantik der Rede: Kontexttheorie - Modalwörter - Konditionalsätze*, Scriptor, Königstein/Taunus.

Kratzer, Angelika (1979), 'Conditional Necessity and Possibility', in Rainer Bäuerle, Urs Egli, and Arnim von Stechow (eds.), *Semantics from Different Points of View,* Springer, Berlin, 117-147.

Kratzer, Angelika (1981a), 'Blurred Conditionals', in Wolfgang Klein and Willem Levelt (eds.), *Crossing the Boundaries in Linguistics. Studies Presented to Manfred Bierwisch,* Reidel, Dordrecht, 201-210.

Kratzer, Angelika (1981b), 'The Notional Category of Modality', in H.J. Eikmeyer and H. Rieser (eds.), *Words, Worlds, and Contexts. New Approaches in Word Semantics,* de Gruyter, Berlin, 38-74.

Kratzer, Angelika (1981c), 'Partition and Revision: The Semantics of Counterfactuals', *Journal of Philosophical Logic,* **10**, 201-216.

Kratzer, Angelika (1986), 'Conditionals', *Chicago Linguistics Society*, **22**, 1-15.

Kratzer, Angelika (1991), 'Modality', in Arnim von Stechow and Dieter Wunderlich (eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, Walter de Gruyter, Berlin, 639-650.

Ladusaw, Bill (1979), *Polarity Sensitivity as Inherent Scope Relations*, Ph.D. Dissertation, University of Texas at Austin.

Landman, Fred (1986), *Towards a Theory of Information. The Status of Partial Objects in Semantics.*, Foris, Dordrecht.

Lewis, David (1979a), 'A Problem with Permission', in Esa Saarinen, Risto Hilpinen, Ilkka Niiniluoto, and Merril Provence Hintikka (eds.), *Essays in Honour of Jaako Hintikka: On the Occasion of His Fiftieth Birthday on January 12, 1979*, Reidel, Dordrecht, 163-175.

Lewis, David (1979b), 'Scorekeeping in a Language Game', *Journal of Philosophical Logic*, **8**, 339-359.

Lewis, David (1996), 'Elusive Knowledge', *Australasian Journal of Philosophy*, **74**, 549-579.

Lowe, E.J. (1990), 'Conditionals, Context, and Transitivity', *Analysis*, **50**, 80-87.

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.

Muskens, Reinhard, Johan van Bethem, and Albert Visser (1997), 'Dynamics', in Johan van Benthem and Alice ter Meulen (eds.), *Handbook of Logic and Language*, Elsevier, Amsterdam, 587-648.

Nute, Donald (1984), 'Conditional Logic', in Dov Gabbay and Franz Guenthner (eds.), *Handbook of Philosophical Logic. Volume II*, Reidel, Dordrecht, 387-439.

Partee, Barbara (1993), 'On the 'Scope of Negation' and Polarity Sensitivity', in Eva Hajicova (ed.), *Functional Approaches to Language Description.*

Roberts, Craige (1986), *Modal Subordination, Anaphora, and Distributivity*, Ph.D. Dissertation, University of Massachusetts at Amherst.

Roberts, Craige (1989), 'Modal Subordination and Pronominal Anaphora in Discourse', *Linguistics and Philosophy,* **12**, 683-721.

Roberts, Craige (1995), 'Domain Restriction in Dynamic Semantics', in Emmon Bach, Eloise Jelinek, Angelika Kratzer, and Barbara Partee (eds.), *Quantification in Natural Languages,* Kluwer, Dordrecht, 661-700.

Stalnaker, Robert (1972), 'Pragmatics', in Donald Davidson and Gilbert Harman (eds.), *Semantics of Natural Language,* Reidel, Dordrecht, 380-397.

Stalnaker, Robert (1973), 'Presuppositions', *Journal of Philosophical Logic,* **2**, 447-457.

Stalnaker, Robert (1974), 'Pragmatic Presuppositions', in Milton Munitz and Peter Unger (eds.), *Semantics and Philosophy,* New York University Press, New York, 197-213.

Stalnaker, Robert (1975), 'Indicative Conditionals', *Philosophia,* **5**, 269-286.

Stalnaker, Robert (1978), 'Assertion', in Peter Cole (ed.), *Syntax and Semantics* **9**, Academic Press, New York, 315-322.

Stalnaker, Robert (1988), 'Belief Attribution and Context', in Robert Grimm and Daniel Merrill (eds.), *Contents of Thought,* University of Arizona Press, Tucson, 140-156.

Stalnaker, Robert (1998), 'On the Representation of Context', *Journal of Logic, Language, and Information* **7**.

Stone, Matthew (1997), 'The Anaphoric Parallel between Modality and Tense', *IRCS Report,* **97**, University of Pennsylvania.

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

Veltman, Frank (1981), 'Data Semantics', in Jeroen Groenendijk, Theo Janssen, and Martin Stokhof (eds.), *Formal Methods in the Study of Language,* Mathematical Centre, Amsterdam,

Veltman, Frank (1985), *Logics for Conditionals*, Ph.D. Dissertation, University of Amsterdam.

Veltman, Frank (1986), 'Data Semantics and the Pragmatics of Indicative Conditionals', in Elizabeth Traugott, Alice ter Meulen, Judy Reilly, and Charles Ferguson (eds.), *On Conditionals*, Cambridge University Press, Cambridge, 147-168.

Veltman, Frank (1996), 'Defaults in Update Semantics', *Journal of Philosophical Logic,* **25**, 221-261.

Warmbrod, Ken (1981a), 'Counterfactuals and Substitution of Equivalent Antecedents', *Journal of Philosophical Logic,* **10**, 267-289.

Warmbrod, Ken (1981b), 'An Indexical Theory of Conditionals', *Dialogue: Canadian Philosophical Review,* **20**, 644-664.

Warmbrod, Ken (1983), 'Epistemic Conditionals', *Pacific Philosophical Quarterly,* 249-265.

Wright, Crispin (1983), 'Keeping Track of Nozick', *Analysis,* **43**, 134-140.