How To Derive Conveyed Meanings

Michael Morreau

Philosophy Department & UMIACS,
University of Maryland at College Park,
College Park, MD 20742-7615,
U.S.A.

Abstract

Below I will analyze defeasible inferences underlying the interpretation of conversational implicatures and presuppositions; a nonmonotonic consequence notion will lead from premises including pragmatic generalizations to conclusions about how those sentences are best interpreted.·

1 Introduction

Suppose you have said something in my presence. Now given our common knowledge of the norms of language use and given reasonable assumptions of mine about your dispositions and cognitive states — given all this, what am I entitled to suppose you have conveyed? And what kind of reasoning allows me to derive what is conveyed from the fact of your utterances? These are the questions to which, borrowing from the philosophy of language and the Artificial Intelligence literature of nonmonotonic reasoning, I will offer an incomplete answer.

Utterances convey meanings beyond their conventional content. I will use the term *conveyed meaning* in a sense which subsumes the conventional consequences of utterances, their conversational implicatures and their "accommodated" presuppositions. Their metaphoric implications and more could be included too, but I won't get to that here; I will focus on what is conveyed through the "accommodation" of presuppositions (in particular the existential presuppositions of definite descriptions) and on what is conveyed by conversational implicatures (in particular the quantity presuppositions of indicative conditionals). I will look too at what happens when these two kinds of conveyed meanings come into conflict.

The following examples, which will be analyzed in detail, illustrate some of these presuppositions, implicatures and conflicts.

Suppose for reasons of your own you say out of the blue, "the cat is not on the mat." And, for the meantime, you say nothing else. Then you convey, among other things, that there is a cat, and that there is a mat. The "accommodation" of the existential presuppositions of definite descriptions has been supposed to play a part in this; here I will suppose simply that speakers of English know that uttering a definite description or another "presuppositional trigger" typically has the effect of conveying the presupposition, without going into how and why it has this effect.

This is not a typical case, though, since you continue with the explanation: "there isn't a cat." Now your two utterances taken together will still convey that there is a mat, but no longer that there is a cat. This illustrates the *defeasibility* of pragmatic reasoning. In the analysis sketched here defeasibility emerges from the nonmonotonic reasoning which, I claim, links premises including pragmatic generalizations and the fact that certain sentences have been uttered to conclusions about what they convey.

Suppose, to take another kind of example, you utter a conditional in the indicative mood: "if my cheque has arrived then I'll pay you back today". I will again be able to draw defeasible conclusions about what this conveys: typically, that you are in a position to assert neither the consequent, "I'll pay you back today", nor the antecedent, "my cheque has arrived" (which together with the conditional itself entails the consequent). Such inferences are standardly treated as conversational implicatures of Grice's conversational maxim of *quantity* and that is how I will treat them here, too.

· Research supported in part by the NSF and ARL.

Interestingly, in cases where there is a conflict, the quantity implicatures of indicative conditionals override the existential presuppositions of definite descriptions occurring in them. If you say "the cat is on the mat if she's in the house" there is no such conflict and your utterance conveys both that there is a cat and that you are unsure whether she is inside. If on the other hand you say "the cat is on the mat, if there is a cat" there is a conflict between the existential presupposition, that there is a cat, and the conversational implicature, that you are unsure whether there is a cat, since you can hardly be supposed to convey both of these things. It seems that in cases like this the conversational implicature wins out: you convey only that you are unsure whether or not there is a cat.

These are the only examples I will consider here. In analyzing them I have tried to keep in mind many, many other cases which a successful theory must account for too. Dealing with those other examples will engage the cogs and wheels which here sometimes turn without doing much useful work.

2 What Knowledge Underlies Interpretation?

The derivation of conveyed meanings is supported by knowledge some of which is broadly speaking linguistic, though much of it is not. In order to express this knowledge precisely it is helpful to introduce, besides the natural language of the utterances, two formal languages. I assume a natural language $\mathcal{L}_N$, a formal language $\mathcal{L}_{LF}$ (here a language of first-order modal logic) in which to write the logical forms of the expressions of $\mathcal{L}_N$ and a language $\mathcal{L}$ in which to formalize the pragmatics of $\mathcal{L}_N$. I assume $\mathcal{L}_{LF}$ is a sublanguage of $\mathcal{L}$. To express what we know about the semantics and pragmatics of $\mathcal{L}_N$, $\mathcal{L}$ contains terms which refer to expressions in those other two languages: for any English sentence σ, "σ" is an individual term of $\mathcal{L}$. And for every $\mathcal{L}_{LF}$ sentence φ, "φ" is an individual term of $\mathcal{L}$. (Goedel numbering is one way of introducing such terms.) Since $\mathcal{L}_{LF}$ is a sublanguage of $\mathcal{L}$, this latter stipulation means that $\mathcal{L}$ has individual terms which must be interpreted as (some of) its own formulas; as far as I can see at the moment this does not introduce paradoxes.

A possible-worlds model for $\mathcal{L}$ has in its domain $\mathcal{D}$ the sentences of $\mathcal{L}_N$ and those of $\mathcal{L}_{LF}$, along with some speakers of $\mathcal{L}_N$ and some things for them to talk about. (We need a cat, and a mat.) Talk about these things is captured in relations on $\mathcal{D}$ and between possible worlds which are suitable for interpreting, in addition to the vocabulary of $\mathcal{L}_{LF}$, the following relation symbols and modal operators. For each speaker s there is a monadic predicate utt_s; for any sentence σ of English, the intended interpretation of utt_s"σ" is that s has uttered (some token of) σ. There is a binary relation lf between (names of) English sentence types and (names of) their logical forms. lf("σ","φ") expresses that "φ" is the logical form of "σ". There is a binary relation $<_i$ between (names of) logical forms; "φ"$<_i$"ψ" means, informally speaking, that an utterance whose logical form is "ψ" would be more informative than one with logical form "φ". For each speaker s and sentence φ of $\mathcal{L}_{LF}$ I assume modal operators $cons_s$ and bel_s; $cons_s(\varphi)$ means that the speaker has conveyed φ; $bel_s\varphi$, that the speaker believes φ. I assume also a modal operator rel; $rel\varphi$ is intended to express that φ is relevant to the dialogue currently underway.[1]

Finally I assume that $\mathcal{L}$ contains a weak modal conditional > with which to express generalizations about how speakers normally use language to convey meanings. The intended interpretation of $\varphi>\psi$ is that if φ, then normally ψ. This conditional is interpreted in the manner of Stalnaker [1968] or Lewis [1973] using the device of possible-worlds selection functions. Its truth conditions are completely standard except for the fact that the modal constraint *centering* (Chellas [1980] calls it *mp*) is not imposed on worlds-selection functions. Such weak conditionals have been used to express pragmatic generalizations before, by Lascarides and Asher [1993] for example. What I will say about presuppositions and conversational implicatures dovetails with and complements their account of the interpretation of temporal and rhetorical relations left implicit in texts.

I can now formalize representative examples of pragmatic generalizations and other information which enables the derivation of conveyed meanings. First, I assume that the interpreter of $\mathcal{L}$ (which I will now assume to be English) knows the conventional content of its sentences. That is, I assume the interpreter knows certain facts about the logical forms of English sentences. Representative examples are

[1] Of course *rel* ought to be given further analysis, perhaps in terms of counterfactuals about the interpreter's epistemic state and the goals of the dialogue. Roughly speaking, a proposition is relevant at a given point in a dialogue if the hearer does not already believe it and if the goals of the dialogue would be furthered were it to be conveyed to him. I can't go into this here.

lf("the cat does not exist", "$\neg\exists x\, cat\, x$"), and

lf("the cat is on the mat if there is a cat", "$\exists x\, cat\, x \Rightarrow on(\mathrm{THE}\, cat, \mathrm{THE}\, mat)$").[2]

That such knowledge should be represented in the way I have just suggested is very simplistic in several ways; for one thing, as linguists since Chomsky have made clear, the fact that we can understand sentences we've never heard before makes in inconceivable that such knowledge is not derived somehow from underlying knowledge of syntax. For another, anaphoric relations between different utterances make the notion that logical forms are formulas of my $\mathcal{L}_{LF}$ exceedingly implausible; alternative representations which would be better include Kamp's *Discourse Representation Structures*, presented in Kamp&Reyle [1994]. This having been said, these oversimplifications do not amount to objections to the approach I am suggesting here, so much as directions in which there is every reason to believe it can be extended in the future.

Concerning the generalizations underlying interpretation, some of them are not linguistic at all but concern the cognitive states of the participants in a dialogue. To begin with, our beliefs tend not to be inconsistent:

CONSISTENCY OF BELIEFS: $\top > \neg bel_S \bot$

That we should know any such thing is implausible and again an oversimplification is responsible. What is required for the pragmatic reasoning formalized below is that speakers typically do not simultaneously entertain two beliefs of which they are aware, and which are obviously incompatible. Such an assumption is a lot more plausible but it is indistinguishable from the generalization stated above if belief is modelled in the simplistic way I've chosen here, as a (nonalethic) modal operator along the lines of . Recent work in Artificial Intelligence on the notion of *awareness* will hopefully give rise to a plausible alternative to my *bel*.

Other generalizations are pragmatic insofar as they derive from the assumption that speakers are governed by Gricean maxims of conversation (and a few others which I will get to). A principle related to Grice's maxim of *quality*, firstly, requires that we convey only what we believe to be true. A speaker who accepts this maxim will normally convey only propositions which he believes:

QUALITY: $con_S\varphi > bel_S\varphi$.

Exceptions to this generalization include lies and other utterances which convey things the speaker fails to believe — which he believes, in the case of lies, to be false. Not all such utterances are insincere though, or intended to mislead; they include guesses, for example. In fact this principle is stronger than Grice's maxim of quality, which required only that speakers avoid utterances they take to be false.

Grice's maxim of *relevance* requires us to convey only what we take to be relevant:

RELEVANCE: $con_S\varphi > bel_S rel\, \varphi$.

Exceptions to the rule that conveyed meanings typically are taken to be relevant include cases in which a speaker moves the conversational goals. Thus for example while waiting for a taxi one might interrupt a trivial converation about some or other topic by announcing that the taxi has arrived. (This example makes clear that the notion of relevance which I intend is a highly context-sensitive one: of course the fact that the taxi has arrived is relevant in a broad sense; that is why it is not considered rude to interrupt with the announcement. But in general it is not relevant to the ongoing conversation, which might as well be about gardening. None of this context sensitivity is present in my analysis, though, so here too there is scope for development of the notion of a conversational goal.)

Finally Grice's maxim of *quantity* requires speakers to be as informative as possible, within limits imposed by quality and relevance. I formalize this by supposing that a speaker normally will prefer a less informative to a more informative utterance only if he cannot assert the more informative thing, either because he does not believe it, or because he does not think it relevant.

QUANTITY: $con_S\varphi \wedge bel_S(\text{"}\varphi\text{"} \lessdot \text{"}\psi\text{"}) \quad > \quad \neg(bel_S\psi \wedge bel_S rel\, \psi)$.

Other pragmatic generalizations which as far as I can see do not derive from broadly Gricean maxims include the law that utterances tend to convey their conventional content:

CONVENTIONAL CONTENT: $utt_S\text{"}\sigma\text{"} \wedge lf(\text{"}\sigma\text{"},\varphi) \quad > \quad con_S\varphi$.

Exceptions to this generalization include utterances which are sarcastic, polite, ironic, metaphorical or tender.

Finally, the following principle expresses the notion that anything which typically holds when a sentence σ is uttered (under any given conditions χ) is typically conveyed by such an utterance, provided it is relevant:

[2] $\Rightarrow$ is a binary connective representing the indicative conditional; THE a definite-description operator which turns a monadic predicate into an individual term.

SIGNIFICATION:

$$\begin{array}{c} rel\ \varphi \quad \wedge \quad \chi \wedge utt_S\text{"}\sigma\text{"} \quad \wedge \quad (\chi \wedge utt_S\text{"}\sigma\text{"}) > \varphi \\ > \\ con_S\varphi \end{array}$$

I expect that some such principle is responsible for the sense of paradox engendered by G.E. Moore's famous utterance: "the cat is on the mat, and I don't believe it". By conventional content and quality the first part of this utterance, "the cat is on the mat", is normally uttered in a context in which the speaker *believes* that the cat is on the mat. From this and the fact that "the cat is on the mat" has in fact been uttered will follow, with signification, that the speaker in uttering these words normally conveys not only that the cat is on the mat, but also that he believes this to be so. Now this latter thing of course directly contradicts the conventional content of the second part of the utterance, "...I don't believe it [i.e. that the cat is on the mat]".

In this illustration (which can be given a rigourous treatment once a notion of nonmonotonic reasoning using > has been introduced), signification was called on to convey certain propositions about the mental state of the speaker. But I have stated the principle more generally than this; it asserts that *any* relevant meaning φ is conveyed by an utterance even if it is associated with that utterances, say, by laws of nature, and does not arise from the conventional meaning of the sentences uttered and pragmatic laws. In fact I will make use of this scheme only with φ instantiated as a sentence of the form $bel_S\psi$ or of the form $\neg bel_S\psi$. It seems to me that this principle is plausible for other propositions φ too, though, especially propositions of which the speaker can be expected to be aware of, so I've stated the principle as generally as possible. Whether it is too permissive has to be seen.

Finally I assume that merely uttering certain lexical items, including "presuppositional triggers" contributes to conveyed meaning. For just one example, the following generalization concerns definite descriptions which, I suggest, tend to convey that there are things fitting those descriptions:

DEFINITE DESCRIPTIONS: $utt_S(\text{"...}the\,\Pi\text{.."}) \;>\; con_S\exists x\Pi$

Here Π is a predicate expression of English but doubles as its formal representation. Similar generalizations can be stated for other presuppositional triggers listed by Levinson [1983]: factive verbs, implicative verbs, change of state verbs, iteratives, verbs of judging, temporal clauses, cleft sentences, comparisons, contrasts and all of the rest.

The pragmatic and other generalization schemes just described make up a background theory against which, in the following, the interpretation of a speaker's utterances will proceed. Call the set of all of their instantiations BT, and let σ be some or other sentence of English. In the following section I will introduce a nonmonotonic consequence notion by means of which it will be possible to derive, from BT and *utterance facts* of the form $utt_S(\text{"}\sigma\text{"})$, conclusions of the form $con_S\varphi$. Meanings φ such that $con_S\varphi$ can be derived from BT and $utt_S(\text{"}\sigma\text{"})$ are the meanings which are conveyed by the speaker's uttering σ.

3 Nonmonotonic Reasoning

Now I introduce a notion of nonmonotonic inference which enables the consequents of these conditionals to be detached in cases where this does not introduce inconsistency.

Conditional logic has been used before as the basis for nonmonotonic reasoning by, among others, Delgrande [1988], Asher&Morreau [1991], and Boutilier [1992]. The form of inference which I will use here is different from all of these; it does not face the problem with "irrelevant" premises which Delgrande and Botilier ran up against and it is from a conceptual and technical point of view not as dismayingly complicated as the notion of "commonsense entailment" introduced by Asher and Morreau. In fact, as a later theorem will show, the inference notion amounts to a trivial fragment of (a slight generalization of) Reiter's Default Logic (the defaults are understood to be defined $\mathcal{L}$, not on classical logic as they are in Reiter's original [1980] presentation.)

The informal idea is that premises including BT and $utt_S(\text{"}\sigma\text{"})$ can be strengthened by adding, as default assumptions, as many instances of *modus ponens*, $\varphi>\psi \rightarrow \varphi\rightarrow\psi$, as is possible without introducing inconsistency. This last scheme I will abbreviate as $\varphi \overset{>}{/\!\!\rightarrow} \psi$. I will also consider examples in which the default assumptions are of the form $(\varphi>\psi \wedge \psi>\chi) \rightarrow (\varphi>\chi)$. The effect of adding these default assumptions is that defeasible forms of the argument forms *modus ponens* and *hypothetical syllogism* become available (neither argument scheme is logically valid):

MODUS PONENS:

$$\begin{array}{l} \varphi \\ \underline{\varphi > \psi} \\ \therefore\ \psi \end{array}$$

HYPOTHETICAL SYLLOGISM:

$$\begin{array}{l} \varphi > \psi \\ \underline{\psi > \chi} \\ \therefore\ \varphi > \chi \end{array}$$

One basic concept is the *focal vector*, which is just a vector of (sets of) default assumptions relative to which nonmonotonic inference will be defined. It plays very much the same role here as Reiter's sets of defaults do, in Default Logic:

DEFINITION: $\mathcal{G} = <\mathcal{G}_1, \mathcal{G}_2, \ldots \mathcal{G}_k>$ is a *focal vector* just in case each $\mathcal{G}_i$ is a set of sentences.

A focal vector should be thought of as a stock of potential assumptions: assumptions which can be added to what is known in any given case where an utterance is to be interpreted. These assumptions come with different priorities: assumptions in $\mathcal{G}_i$ have higher priority than those in $\mathcal{G}_j$, if i<j. In cases of conflict between potential assumptions, what will be assumed and what will not is determined by this priority order, since an assumption of lower priority will never be made if it conflicts with an assumption of higher priority which could have been made in its place. Assumptions in $\mathcal{G}_1$ will take precedence in conflicts with those in $\mathcal{G}_2$, $\mathcal{G}_3$,...; assumptions in $\mathcal{G}_2$ will take precedence in conflicts with those in $\mathcal{G}_3$, $\mathcal{G}_4$,... and so on. Below, $\mathcal{G}$ will be a vector $<\mathcal{G}_1, \mathcal{G}_2, \ldots \mathcal{G}_k>$ of (sets of) instances of *modus ponens* whose effect, when assumed, is to enable conclusions to be drawn about what is conveyed in actual utterance situations.

An example illustrates assumption-making in which higher-priority assumptions concerning the conventional content of an utterance override lower-priority assumptions based on the speaker's choice of definite-description syntax. Consider the following two utterance situations:

EXAMPLE:
(I) A speaker says "the cat is not on the mat" (and nothing else).
(II) A speaker says "the cat does not exist" (and nothing else).

In the first case the speaker conveys that there is a cat, which is not on the mat. In the second case the speaker conveys that there isn't a cat. Let $\mathcal{G}$ be the vector $<\mathcal{G}_1,\mathcal{G}_2>$, of which $\mathcal{G}_1$ contains just the following five instances of *modus ponens* (all can be gotten from BT in a way which will become clear as we go). Adding the assumptions in $\mathcal{G}$ to the utterance facts as described in (I) and (II) will, relative to BT, enable the interpreter to derive appropriate conclusions about what the speaker has conveyed in each of these situations.

First, $\mathcal{G}_1$ contains $\top \overset{>}{/\!\!\rightarrow} \neg bel_s\bot$. Relative to a background theory including CONSISTENCY OF BELIEFS, to assume this sentence is of course simply to assume $\neg bel_s\bot$, which expresses that the speaker's beliefs are consistent.

Second, $\mathcal{G}_1$ contains the following two sentences:

$$cons_s(\neg\exists x\, cat\, x) \overset{>}{/\!\!\rightarrow} bel_s(\neg\exists x\, cat\, x)$$
$$cons_s(\exists x\, cat\, x) \overset{>}{/\!\!\rightarrow} bel_s(\exists x\, cat\, x).$$

Relative to a background theory including QUALITY, and in a context in which the speaker has conveyed that there is (not) a cat, to assume these sentences is simply to assume that the speaker believes what he has conveyed.

Finally, $\mathcal{G}_1$ contains the following two instances of *modus ponens* which, so to speak, cash out the generalization CONVENTIONAL CONTENT. The first is

$$\begin{array}{c} utt_s\text{"the cat is not on the mat"} \wedge \\ lf(\text{"the cat is not on the mat"}, \text{"}\neg on(\text{THE}cat, \text{THE}mat)\text{"}) \\ \overset{>}{/\!\!\rightarrow} \\ cons_s(\neg on(\text{THE}cat, \text{THE}mat)) \end{array}$$

This sentence expresses the following thing: if an utterance of "the cat is not on the mat" *normally* conveys the conventional content of this sentence, namely that the cat is not on the mat, and if in addition this sentence has been uttered, then the speaker has conveyed that the cat is not on the mat. The other sentence in $\mathcal{G}_1$ expresses the analogous thing about the sentence "the cat does not exist" and *its* conventional content:

$$\begin{gathered} utt_S\text{"the cat does not exist"} \wedge lf(\text{"the cat does not exist"}, \text{"}\neg\exists x\, cat\, x\text{"}) \\ >\!\!/\!\!\!\rightarrow \\ cons_S(\neg\exists x\, cat\, x) \end{gathered}$$

$\mathcal{G}_2$ is chosen to include sentences which cash out the generalization DEFINITE DESCRIPTIONS in much the same way. (It is significant that they appear in $\mathcal{G}_2$ instead of $\mathcal{G}_1$. This is how the interpretation process incorporates an empirical hypothesis which I now put to you: that assumptions deriving from the speaker's choice of syntax have lower priority than assumptions deriving from his choice of conventional content, assumptions about the consistency of the speaker's beliefs, and so on.) The first sentence in $\mathcal{G}_2$ is,

$$utt_S\text{"the cat is not on the mat"} \quad >\!\!/\!\!\!\rightarrow \quad cons_S\exists x\, cat\, x.$$

It expresses the following thing: if an utterance of "the cat is not on the mat" normally conveys that there is a cat, and if in addition this sentence has in fact been uttered, then the speaker has in fact conveyed that there is a cat.

The second sentence in $\mathcal{G}_2$ does the same thing for the sentence "the cat does not exist". This is an interesting case because when this sentence is uttered its conventional content and the use of definite-description syntax tend to pull the interpreter of the utterance in different directions:

$$utt_S\text{"the cat does not exist"} \quad >\!\!/\!\!\!\rightarrow \quad cons_S\exists x\, cat\, x.$$

This sentence expresses that if an utterance of "the cat does not exist" normally conveys that there is a cat, and if in addition this sentence has in fact been uttered, then the speaker has in fact conveyed that there is a cat. We will see that because this assumption has lower priority than the sentence dealing with the conventional content of "the cat does not exist", the effect of uttering this sentence is simply to convey that there is no cat.

I am unhappy that the procedure I've just followed in deriving defaults from BT (this procedure will be stated explicity below) should lead to the inclusion in our focal vector of an assumption as implausible as this one! Its presence doesn't seem to do much damage insofar as it does not lead us to derive counterintuitive conveyed meanings, but it is a blight on my account. The problem seems to be that generalizations should not be treated as I have treated them, as schemes or universally quantified sentences. It is true I think that a definite description normally is uttered in contexts in which the existence is conveyed of a unique salient thing satisfying the description. But this should not entail that, say, "the cat does not exist" is normally uttered in contexts in which the existence is conveyed of a unique salient thing satisfying the description (a cat in this case). This question and its answer are discussed in a little more detail in my paper "Allowed Arguments", which appears in this IJCAI.

Now the picture of the derivation of conveyed meanings which I will develop here is in outline as follows.

Take situation (I) above: someone utters "the cat is not on the mat" in the presence of an interpreter. The interpreter observes this utterance and his observation leads him to adopt, as a premise of his interpretative reasoning, utt_S"the cat is not on the mat". Also, syntactic analysis contributes the premise lf("the cat is not on the mat", "$\neg on$(THEcat, THEmat)"). Now to these premises and to the pragmatic and other generalizations in BT we add as many of the above assumptions as we can without introducing inconsistencies. We start with the assumptions in $\mathcal{G}_1$ and find that all five can be added. Then we go on to add as much of $\mathcal{G}_2$ as we can without introducing inconsistencies, finding that all of $\mathcal{G}_2$ can be added, too. (Of course consistency must be demonstrated, say, with a model construction.) So the assumptions we are able to make are just $\mathcal{G}_1\cup\mathcal{G}_2$. Now it can be verified that relative to BT, our premises and assumptions entail, among other things, both $cons_S\exists x\, cat\, x$ and $cons_S\neg on$(THEcat, THEmat). That is,

$$\begin{gathered} \{utt_S\text{"the cat is not on the mat"}, lf(\text{"the cat is not on the mat"}, \text{"}\neg on(\text{THE}cat, \text{THE}mat)\text{"}\} \\ \text{BT} \cup \mathcal{G}_1 \cup \mathcal{G}_2 \\ \models \\ cons_S(\exists x\, cat\, x) \\ cons_S(\neg on(\text{THE}cat, \text{THE}mat)) \end{gathered}$$

In this way the interpreter is able to derive both $\exists x\, cat\ x$ and $\neg on(\text{THE}cat, \text{THE}mat)$ as conveyed meanings in *situation (I)*.

Take situation (II) above: the speaker utters "the cat does not exist" in the presence of the interpreter. This leads the interpreter once again via syntactic processing to premises $\{utt_S$"the cat does not exist", lf("the cat does not exist", "$\neg\exists x\, cat\ x$")$\}$. The assumption-making process begins once again with the higher-priority assumptions in $\mathcal{G}_1$, finds that they can all be added without loss of consistency, and moves on to see how much of $\mathcal{G}_2$ can be added. Now note that

$$\begin{gathered}\{utt_S\text{"the cat does not exist"}, lf(\text{"the cat does not exist"}, \text{"}\neg\exists x\, cat\ x\text{"}\}\\ \text{BT}\cup\mathcal{G}_1\\ \models\\ con_S(\neg\exists x\, cat\ x)\end{gathered}$$

The main player in the demonstration that this is so is the sentence

$$\begin{gathered}utt_S\text{"the cat does not exist"}\wedge lf(\text{"the cat does not exist"}, \text{"}\neg\exists x\, cat\ x\text{"})\\ >\!/\!\!\rightarrow\\ con_S(\neg\exists x\, cat\ x)\end{gathered}$$

which with great foresight was included in $\mathcal{G}_1$. Other elements of $\mathcal{G}_1$ enable us to derive as well the following three things:

$$\begin{gathered}\neg bel_S\bot\\ con_S(\neg\exists x\, cat\ x) \rightarrow bel_S(\neg\exists x\, cat\ x)\\ con_S(\exists x\, cat\ x) \rightarrow bel_S(\exists x\, cat\ x)\end{gathered}$$

Putting all of these premises together (and assuming — with less than full plausibility — that the logic of *bel* is such that believing contradictories entails believing $\bot$) we have

$$\begin{gathered}\{utt_S\text{"the cat does not exist"}, lf(\text{"the cat does not exist"}, \text{"}\neg\exists x\, cat\ x\text{"}\}\\ \text{BT}\cup\mathcal{G}_1\\ \models\\ \neg con_S(\exists x\, cat\ x)\end{gathered}$$

The effect of this is that the following sentence, which is in $\mathcal{G}_2$, cannot consistently be added to these premises.

$$utt_S(\text{"the cat does not exist"}) \ >\!/\!\!\rightarrow\ con_S(\exists x\, cat\ x)$$

The other element of $\mathcal{G}_2$ can be though. So in utterance situation (II) the result of assuming as much as possible while respecting priorities is just

$$\begin{gathered}\{utt_S\text{"the cat does not exist"}, lf(\text{"the cat does not exist"}, \text{"}\neg\exists x\, cat\ x\text{"}\}\\ \text{BT}\cup\mathcal{G}_1\cup\{utt_S(\text{"the cat is not on the mat"}) \ >\!/\!\!\rightarrow\ con_S(\exists x\, cat\ x)\}\end{gathered}$$

We have seen that this set entails $con_S(\neg\exists x\, cat\ x)$ but (by its consistency) does not entail $con_S(\exists x\, cat\ x)$. Thus by this assumption-making the interpreter is able to derive, from the observed facts of utterance situation (II) and his background theory of interpretation, that the speaker has conveyed $\neg\exists x\, cat\ x$. But other than in utterance situation (I), the speaker's choice of definite-description syntax does not lead the interpreter to suppose that he has also conveyed $\exists x\, cat\ x$.

The previous informal discussion of assumption-making illustrates the following precise notions. First a definition. Letting $\mathcal{G} = \langle\mathcal{G}_1, \mathcal{G}_2, \ldots \mathcal{G}_k\rangle$ be a focal vector and letting $\mathcal{H}$ be a set of assumptions disjoint from every $\mathcal{G}_i$, $\mathcal{G}|\mathcal{H}$ is a useful notation for the focal vector $\langle\mathcal{G}_1, \mathcal{G}_2, \ldots \mathcal{G}_k, \mathcal{H}\rangle$. Definitions and proofs using this notation can proceed by induction on the length of focal vectors. For a trivial start, a focal vector $\mathcal{F}$ can be "flattened" into $\mathcal{F}^{\cup}$ as follows: $\langle\rangle^{\cup} = \{\}$; $(\mathcal{G}|\mathcal{H})^{\cup} = \mathcal{G}^{\cup} \cup \mathcal{H}$ The following notion captures the idea of a set Ω of assumptions, taken from $\mathcal{F}$ in order of their priority and consistent with premises Γ, than which no other such set is more inclusive:

DEFINITION: Ω *is maximal* Γ*–satisfiable within* $\mathcal{I}$ is defined by induction on the length of $\mathcal{I}$:

Base step: { } is maximal Γ–satisfiable within <>.

Induction step: Ω is maximal Γ-satisfiable within $\mathcal{G}|\mathcal{H}$ if

i. $\Omega \subseteq (\mathcal{G}|\mathcal{H})^{\cup}$;

ii. $\Omega \cap \mathcal{G}^{\cup}$ is maximal Γ-satisfiable within $\mathcal{G}$

iii. $\Gamma \cup \Omega \not\models \perp$, and if $\Omega \subseteq \Omega^* \subseteq (\mathcal{G}|\mathcal{H})^{\cup}$ and $\Gamma \cup \Omega^* \not\models \perp$, then $\Omega = \Omega^*$.

DEFINITION: *prioritized allowed entailment* $\Gamma \models_{\mathcal{I}} \varphi$ means that for every Ω which is maximal Γ-satisfiable within $\mathcal{I}$, $\Gamma\Omega \models \varphi$.

The following continuation of the earlier examples shows how $\models_{\mathcal{I}}$ is used to formalize the derivation of conveyed meanings from premises including pragmatic generalizations and the fact that given sentences have been uttered.

CONINUATION OF EXAMPLE: Let $\mathcal{G} = <\mathcal{G}_1, \mathcal{G}_2>$ be as in the earlier informal discussion.

(I) A speaker says "the cat is not on the mat" and nothing else. Let Γ be BT$\cup${utt_s"the cat is not on the mat"}. The earlier discussion is a demonstration that there is a unique $\mathcal{G}$-respecting and maximal Γ-consistent subset of $\mathcal{G}_1 \cup \mathcal{G}_2$, and that is $\mathcal{G}_1 \cup \mathcal{G}_2$ itself. Also, it was shown there that $\Gamma \cup \mathcal{G}_1 \cup \mathcal{G}_2 \models cons_s(\exists x\, cat\, x)$ and also it was shown that $\Gamma \cup \mathcal{G}_1 \cup \mathcal{G}_2 \models cons_s(\neg on(\text{THE}cat, \text{THE}mat))$. Thus uttering "the cat is not on the mat" conveys that there is a cat, which is not on the mat:

BT$\cup${utt_s"the cat is not on the mat"}

$\models_{\mathcal{G}}$

$cons_s(\exists x\, cat\, x)$

$cons_s(\neg on(\text{THE}cat, \text{THE}mat))$

(II) A speaker says "the cat does not exist" and nothing else. Let Γ be BT$\cup${utt_s"the cat does not exist"}. The earlier discussion is a demonstration that there is a unique maximal Γ-consistent set within $\mathcal{G}$, and that is $\mathcal{G}_1 \cup${utt_s("the cat is not on the mat") $>\!/\!\!\rightarrow$ $cons_s(\exists x\, cat\, x)$}.

Call this set Ω. It was shown earlier that $\Gamma \cup \Omega \models cons_s(\neg\exists x\, cat\, x)$, though $\Gamma \cup \Omega \not\models cons_s(\exists x\, cat x)$. Thus uttering "the cat does not exist" conveys that there is no cat, and fails to convey that there is a cat.

BT$\cup${utt_s"the cat does not exist"}

$\models_{\mathcal{G}}$	$\not\models_{\mathcal{G}}$
$cons_s(\neg\exists x\, cat\, x)$	$cons_s(\exists x\, cat x)$

Before going on to analyze some examples though, it may be helpful to compare $\models_{\mathcal{I}}$ with Reiter's [1980] Default Logic. I will show how $\models_{\mathcal{I}}$ can be characterized using a generalization of Reiter's [1980] Default Logic which allows for default rules to have different priorities.

DEFINITION (Reiter): Let φ, ψ and χ be sentences of $\mathcal{L}$. A *default rule* is a rule of inference:

$$\frac{\varphi : \psi}{\chi}$$

It is convenient to write this rule φ:ψ/χ. A *normal* default is a default rule of the form

$$\frac{\varphi : \psi}{\psi}$$

I will write such a rule φ/ψ.

DEFINITION (Reiter): A *(normal) default theory* is a pair (Γ,Δ) where $\mathcal{L}$ is a set of sentences and Δ is a set of (normal) default rules.

This notion can be generalized as follows to allow for a priority order on sets of default rules. Let $\Delta = <\Delta_1, \Delta_2, ... \Delta_k>$ be a finite vector of sets of default rules. Then <Γ,Δ> is a *prioritized default theory*.

DEFINITION (Reiter): An *extension* of (Γ,Δ) is any set $\mathcal{E}$ of $\mathcal{L}$ sentences such that $\mathcal{E} = \cup\mathcal{E}_i$, where each $\mathcal{E}_i$ is defined in terms of $\mathcal{E}$ and (Γ,Δ) as follows:

$\mathcal{E}_0 = \Gamma$,

$\mathcal{E}_{i+1} = \mathrm{Cn}(\mathcal{E}_i) \cup \{\chi$: for some φ and ψ, $\varphi \in \mathcal{E}_i$, $\varphi:\psi/\chi \in \Delta$ and $\neg\psi \notin \mathcal{E}\}$.

Here $\mathrm{Cn}(\mathcal{S}) = \{\varphi: \mathcal{S} \models \varphi\}$. This notion can be recursively extended to prioritized default theories. As for the base clause, we say that $\mathcal{E}$ is an extension of $(\Gamma,\langle\rangle)$ just in case $\mathcal{E} = \mathrm{Cn}(\Gamma)$. For the inductive clause, we say that $\mathcal{E}$ is an extension of $\langle\Gamma,\langle\Delta_1, \Delta_2, \ldots\Delta_{k+1}\rangle\rangle$ just in case there is some extension $\mathcal{E}^*$ of $\langle\Gamma,\langle\Delta_1, \Delta_2, \ldots\Delta_k\rangle\rangle$ such that $\mathcal{E}$ is an extension of $\langle\mathcal{E}^*,\Delta_{k+1}\rangle$ is the original sense of Reiter's.

Now for any sentence φ define Δ_φ, *the default rule corresponding to* φ, to be the normal default T/φ. Note that these are particularly trivial normal default rules. Now generalize to sets of sentences and focal vectors as follows: for any set $\mathcal{F}$ of sentences, $\Delta_{\mathcal{F}} = \{\Delta_\varphi : \varphi \in \mathcal{F}\}$; and for any focal vector $\underline{\mathcal{F}} = \langle\mathcal{F}_1, \mathcal{F}_2, \ldots\mathcal{F}_k\rangle$, $\Delta_{\underline{\mathcal{F}}} = \langle\Delta_{\mathcal{F}_1}, \Delta_{\mathcal{F}_2}, \ldots\Delta_{\mathcal{F}_k}\rangle$. Then the following thing can without much difficulty be proven by induction on the length of $\underline{\mathcal{F}}$: Let $\underline{\mathcal{F}}$ be a focal vector. Then for any Ω which is maximal $\underline{\mathcal{F}}$-respecting Γ-consistent within $\underline{\mathcal{F}}$, $\mathrm{Cn}(\Gamma\cup\Omega)$ is an extension of $\langle\Gamma,\Delta_{\underline{\mathcal{F}}}\rangle$. Conversely, any extension of $\langle\Gamma,\Delta_{\underline{\mathcal{F}}}\rangle$ can be written $\mathrm{Cn}(\Gamma\cup\Omega)$, for some Ω which is maximal $\underline{\mathcal{F}}$-respecting Γ-consistent within $\underline{\mathcal{F}}$. Thus the following equivalence theorem can be stated:

THEOREM: $\Gamma \models_{\underline{\mathcal{F}}} \varphi$ iff for every default extension $\mathcal{E}$ of $\langle\Gamma,\Delta_{\underline{\mathcal{F}}}\rangle$, $\varphi \in \mathcal{E}$.

In "Allowed Arguments" [IJCAI 1995] I show that allowed entailment can also be characterized as what, generalizing John McCarthy's [1980] notion of Minimal Entailment (the model-theoretic pendant of Circumscription) has been called a "preferential" entailment notion. The preference relation is in this characterization an alphabetic preference order on possible worlds determined by $\underline{\mathcal{F}}$. (Thus one world, u, is preferred to another, v, if u satisfies more of $\mathcal{F}_1$ than does v; or if they satisfy the same sentences of $\mathcal{F}_1$ but u satisfies more of $\mathcal{F}_2$ than does v; or if they satisfy the same sentences of $\mathcal{F}_1 \cup \mathcal{F}_2$ but u satisfies more of $\mathcal{F}_3$ than does v; or if....)

The interest of these characterizations here is that interpretation, if it admits an analysis like the one I suggest, is of a kind with forms of nonmonotonic reasoning like Default Logic and Circumscription which are already quite well understood.

4 Back to the Examples

It is now possible to analyze the examples I started out with (and many more besides which I don't have room to treat here.) Let Γ be a background theory containing the pragmatic generalizations and other "knowledge" of section 2 above.

The sentences which appear in some $\mathcal{F}_i$ are said to be *in focus*. Exactly *which* instances of *modus ponens* will be in focus when interpreting any given utterance or text? And what priority do they have? Are there assumptions to be made in interpretation which are not instances of *modus ponens* ? I expect that the factors determining the choice of focus for any given interpretation task will turn out to be many and varied. But for now, and by way of illustration, I can make explicit the strategy which was followed earlier on, in the discussion of example 1. Taking a generalization, say, utt_s("...*the* Π...")$>\mathit{con}_s\exists x\Pi$, an *instantiation* is a sentence which like

utt_s("the king of France is bald")$>\mathit{con}_s\exists x\,\mathit{king\text{-}of\text{-}france}\ x$

substitutes into the generalization, in this case, the name of a sentence of English, and an appropriate predicate expression. A sentence $A{>}/\!\!\rightarrow B$, where $A>B$ instantiates the generalization γ, can be said to *activate* γ. In the examples below I will assume that *all* instances of *modus ponens* are in focus which activate one or other of the above pragmatic generalizations.

Having decided which sentences will appear somewhere in our focal vector $\underline{\mathcal{F}}$ it remains to decide in which $\mathcal{F}_i$ they will appear. In the example I discussed above I supposed that some assumptions have a higher priority than others in pragmatic reasoning. In particular I supposed that assumptions of a general conversational nature, and those concerned with the speaker's choice of conveyed meaning, have higher priority than assumptions concerned with the speaker's choice of syntax. Assume a hierarchy of pragmatic generalizations in which DEFINITE DESCRIPTIONS and other generalizations deriving from presuppositional triggers are ranked lower than QUALITY, QUANTITY and the rest (compare Gazdar's [1979] suggestion). Here, the lower the index i of the $\mathcal{F}_i$ of which $\varphi{>}/\!\!\rightarrow\psi$ is a member, the higher the priority of the assumption ψ. In this way, through these indices, focal vectors can reflect the pragmatic hierarchy: suppose there are pragmatic generalizations γ_1 and γ_2 of which $\varphi_1>\psi_1$ and $\varphi_2>\psi_2$, respectively, are instantiations. And suppose on the pragmatic hierarchy γ_1 is ranked higher than γ_2. And let there be i and j such that $\varphi_1 > / \!\!\rightarrow \psi_1 \in \mathcal{F}_i$ and $\varphi_2 > / \!\!\rightarrow \psi_2 \in \mathcal{F}_j$ (which is to say, suppose we are focusing on these generalizations). Then it is required that i<j.

The claims I make about which conveyed meanings can be derived and which cannot all concern this same choice of the focus of attention except the last one, which introduces sentences other than instances of *modus ponens* into the focus of attention.

the cat isn't on the mat A speaker utters "the cat isn't on the mat" (and nothing else), conveying among other things that there is a cat, and a mat. We have

$$\Gamma,\ utt_S(\text{"the cat isn't on the mat"}) \models_{\mathcal{I}} cons_S(\exists x\, cat\ x) \wedge cons_S(\exists x\, mat\ x)$$

The following demonstration sketch illustrates the general approach. It is not difficult to show with a model construction (and it is not surprising) that Γ, utt_S("the cat is not on the mat") is satisfiable together with *everything* in $\mathcal{I}$. So there is just one Ω which is maximal $\mathcal{I}$-respecting and $\Gamma \cup \{utt_S(\text{"the cat is not on the mat"})\}$-satisfiable within $\mathcal{I}$ and that is $\mathcal{I}^{\cup}$. Now

$$\begin{array}{c}\Gamma,\ utt_S(\text{"the cat isn't on the mat"}) \\ utt_S(\text{"...}the\ \Pi\text{..."}) > / \!\!\rightarrow cons_S \exists x \Pi \\ \models \\ cons_S(\exists x\, cat\ x) \wedge cons_S(\exists x\, mat\ x)\end{array}$$

if the cat isn't on the mat then it's in the kitchen This example illustrates the "projection" of presuppositions into a conditional context. A speaker utters "if the cat is not on the mat then it's in the kitchen" (and nothing else), conveying among other things that there is a cat, a mat, and a kitchen. Now we have:

Γ, utt_S("the cat is not on the mat") $\models_{\mathcal{I}}$ $cons_S(\exists x\, cat\ x) \wedge cons_S(\exists x\, mat\ x) \wedge cons_S(\exists x\, kitchen\ x)$

there isn't a cat A speaker utters two sentences: "The cat is not on the mat" and, by way of explanation, "There isn't a cat." (and nothing else). In so doing he does not convey that there is a cat, but does convey that there is a mat. Let Γ^* stand for $\Gamma \cup \{utt_S(\text{"the cat is not on the mat"}), utt_S(\text{There is no cat"})\}$. Now $cons_S(\neg\exists x\, cat x)$ and $cons_S(\exists x\, mat\ x)$ follow from Γ^* with $\models_{\mathcal{I}}$. $cons_S(\exists x\, cat\ x)$ does not.

the cat is on the mat if there is a cat A speaker utters "the cat is on the mat if there is a cat." Let Γ^* stand for $\Gamma \cup \{utt_S(\text{"the cat is not on the mat if there is a cat"})\}$. Now $cons_S(\neg bel_S \exists x\, cat\ x)$, $cons_S(\neg bel_S \neg\exists x\, cat\ x)$, $cons_S(\neg bel_S\ (on(\text{THE}cat, \text{THE}mat)))$ and $cons_S(\neg bel_S(\neg on(\text{THE}cat, \text{THE}mat)))$ follow with $\models_{\mathcal{I}}$ from Γ^*. The speaker's choice of the indicative conditional conveys his agnosticism about the antecedent and the consequent. The generalization QUALITY plays an important part in the demonstration. On the other hand $cons_S(\exists x\, cat\ x)$ does not follow. This existential presupposition of the expression "the cat" is overridden by the conversational implicature that the speaker is agnostic about the antecedent.

5 Abnormality Predicates and Default Rules

It is important to say why I chose to represent pragmatic generalizations using the modal conditional operator >, and not as first-order sentences involving "abnormality" predicates as in familiar applications of McCarthy's [1980, 1986] Circumscription, or as default rules in the sense of Reiter's [1980] Default Logic. The reason is that the object languages of these theories are too restrictive.

Circumscription is defined within the extensional setting of classical first-order logic and to generalize to more expressive languages is not trivial. (See for example Thomason [1990].) This commitment to extensional logic makes it difficult to see how Circumscription could capture, say, nonmonotonic reasoning involving propositional attitudes, conditionals and the like, all of which figure in pragmatic generalizations. Default Logic in Reiter's original [1980] formulation also takes a classical perspective though in this case that is inessential. Another limitation is however not so easily overcome. The defaults of Reiter's theory are metalinguistic rules, not a part of the object language, so it is not possible to reason *about* them in Default Logic. In particular, default rules cannot be nested into other sentences. But *all* of the pragmatic generalizations I stated hold only under the assumption that the speaker is being cooperative, so that QUALITY for example should really be stated $cooperative(s) \rightarrow (cons_S\varphi > bel_S\varphi)$. For another example, my principle of signification requires that one > can appear within the scope of another >, whereas default rules in Reiter's theory cannot be nested one within the other.

This point that > is not to be thought of as a default rule of inference should of course not be confused with the point expressed in an earlier theorem, that the notion of nonmonotonic inference I have defined on the language of > can be characterized using defaults over this language.

6 Future Work

Priorities include integrating with Lascarides and Asher's work on rhetorical relations, and a comparison with Gazdar [1979].

References

Asher, N. and M. Morreau: 1991, 'Commonsense Entailment: a modal theory of nonmonotonic reasoning.' In J. Mylopoulos and R. Reiter, eds., *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, Morgan Kaufmann, Los Altos, California, pp. 387-392.

Boutilier, C.: 1992, *Conditional Logics for Default Reasoning and Belief Revision.* Technical Report KRR-TR-92-1, Computer Science Department, University of Toronto, Ontario, 1992.

Chellas, B.: 1980, *Modal Logic, and Introduction*, Cambridge University Press, Cambridge.

Delgrande, J.: 1988, 'An Approach to Default Reasoning Based on a First-Order Conditional Logic: revised report.' *Artificial Intelligence 36* (1988), pp. 63-90.

Gazdar, G.: 1979, *Pragmatics: Implicature, Presupposition, and Logical Form.* New York: Academic Press.

Grice, H. P.: 1989, *Studies in the Way of Words.* Harvard.

Kamp&Reyle: 1994, *From Discourse to Logic.* Kluwer.

Lascarides, A., N. Asher.: 1993, 'Temporal Interpretation, Discourse Relations and Commonsense Entailment.' *Linguistics and Philosophy 16*, pp. 437-494.

Lewis, D.: 1973, *Counterfactuals*, Harvard University Press.

McCarthy, J.: 1980, 'Circumscription — a Form of Nonmonotonic Reasoning,' *Artificial Intelligence* 13.

Morreau, M.: 1995, 'Three ways of assumption-making'. Ms. to be submitted to *the Journal of Philosophical Logic.*

Perrault, R.: 1990. 'An Application of Default Logic to Speech Act Theory.' In *Intentions in Communication.* P. Cohen, J. Morgan, and M. Pollack eds., Cambridge, MA: MIT Press, 1990.

Stalnaker, R.: 1968, 'A Theory of Conditionals,' in N. Rescher (*ed.*), *Studies in Logical Theory*, Blackwell, Oxford, 1968, pp. 98-112.

Stalnaker, R.: 1981, 'Indicative Conditionals,' in *Ifs*, W.L. Harper, R. Stalnaker and G. Pearce (*eds.*), D. Reidel Publishing Company, Dordrecht.

Thomason, R.: 1990, 'Propagating Epistemic Coordination through Mutual Defaults I,' In Rohit Parikh, ed., *Proceedings of the third conference on Theoretical Aspects of Reasoning about Knowledge.* Morgan Kaufmann, San Mateo, CA, pp. 29-39.