"He will make you take it back":
A STUDY IN THE PRAGMATICS OF LANGUAGE

Eugene Charniak

1974

Istituto per gli Studi Semantici e Cognitivi
Castagnola, Switzerland

Distribuzione:
Centro di Documentazione
della
Fondazione Dalle Molle per gli studi linguistici e
di comunicazione internazionale
Villa Barbariga
30039 San Pietro di Strà / Italy
About half of the research which went into this paper was done while the author was a research associate at the Artificial Intelligence Laboratory, a Massachusetts Institute of Technology research program supported in part by the Advanced Research Projects Agency of the Department of Defense and monitored by the Office of Naval Research under Contract number N00014-70-A-0362-0003. The latter portion of the research, and actually writing the paper was done at the author's current address:

Istituto per gli studi semantici e cognitivi  
Villa Heleneum  
6976 Castangola, Switzerland

The author thanks both Institutions for their support.
§1 INTRODUCTION

§1.1 What is to be Done

While reading a story, a person fills in the blanks, so to speak, by using the knowledge he has of the world. I am interested in formalizing such "real world knowledge". In particular this paper is concerned with explaining how people understand the last line of the following story.

(1) Today was Jack's birthday. Janet and Penny went to the store. They were going to get presents.

    "I will get a top," said Janet. "Don't do that," said Penny. "Jack has a top. He will make you take it back."\(^2\)

Although people have to be prodded before they realize it, "take it back" does not "mean" taking it to the store for the purpose of exchanging it. Rather, we have knowledge which enables us to make this inference. What this knowledge is, and how we know when to access it, is the subject of this paper.

Even in this seemingly modest goal this paper is only partially successful. What follows is an incomplete analysis.

\(^2\)This example originally appeared in (Charniak '72) where it is an example of a difficult pronoun problem (the "it" in the last sentence). While I became interested in this story because of this pronoun problem, the present paper has developed independently of it.
However, although incomplete, I believe that the research has turned up several interesting points, two of which are worth commenting on now.

First, there are actually two sets of facts, each one by itself sufficient to make the inference in the last line of (1). Roughly, one inference stems from the special significance of "take back" and how it interacts with certain facts about "trading" and "stores". The other set of facts related "stores" to what to do about objects, which, for some reason, are unwanted. While some cases of comparatively low semantic/pragmatic redundancy have been studied (e.g., the use of "but" rather than "and"), this is, to my knowledge, the first example of substantial redundancy to be studied in great detail.

The second interesting result is a small piece of evidence for the intuitive idea that there are different modes of discourse. In particular, it will be argued that there are at least two modes, called conversational and narrative. They are distinguished by the fact that the narrative mode allows statements to be connected to the rest of the discourse by more tenuous (less sure) inferences than would be allowed in conversational mode. We will need this distinction to account for difference in acceptable language between (1) and similar stories where Penny starts telling Janet a "story" about what will happen if she buys Jack a top.

Both of these conclusions are documented in section 2. Section 3 investigates a problem raised in 2, finding a set of rules which would explain why Janet would take the top back. In order to do this for the general case (i.e., both
with and without coercion) we will begin to formalize a bit of naive motivational theory. Section 4 is a more detailed analysis of the "take back" rule presented in 2, and section 5 is a brief conclusion.

The program (scientific, not computer) presented here is to combine the methodology found in transformational linguistics with the goals and assumptions found in artificial intelligence work on language. As for the former, it is assumed that everybody is familiar with the "capture the relevant generalization" type argument. In linguistics, of course, the generalizations are on grammaticality intuitions. Here we will be working instead with intuitions about correct answers to questions. So, given example (1), if we asked "Why doesn't Jack want a top?", "Because he already has one" would be a reasonable answer, but not "I didn't know that he didn't want a top". While the primary source of data is intuitions about answers to questions, other intuitions will be allowed as data, such as "The story makes more sense if X happens than if Y", or "Yes, I understand what the story is saying, but that is not the way one would normally say it."

The theory (or "model", I use the words interchangeable) which accounts for these intuitions can be thought of as a computer program which takes as input a piece of connected discourse (or story) and questions about it. The output consists of answers which people consider "reasonable". If a person can give more than one reasonable answer (ignoring paraphrase) then the model should be able to do the same. If people disagree about what the answer is, then different ver-
sions of the basic model should be capable of producing the
different answers.

§1.2 Assumptions

The assumptions made about the model are fairly standard
in the AI literature. In fact, given how little is known about
language comprehension, the assumptions given here are quite
vague. Nevertheless, it is an empirical issue whether the
adoption of a particular assumption makes it easier or harder
to construct the theory.

The natural language input is transformed into an
appropriate semantic representation. This representa-
tion should be chosen so as to maximize ease of
both input/output and inference.

The data base is made up of assertions, an assertion
being a single predicate with the proper number of
arguments. To put an assertion into the data base
is to assert it on. Assertions have names (n1, n2,...) so that one assertion can refer to another. The
representation will certainly need constants (DOG22
is a particular dog, say the one which belongs to
JACK5) which can be referred to from sentence to
sentence, and will most likely need variables,
quantifiers, and logical connectives, although
whether they will be much like the familiar ones
found in predicate calculus is an open question. 3

Since the discourse could stop at any point and a question asked, the input is precessed on a line basis.

The semantic representation of the story is stored in a data base which is associative in the sense that one can retrieve a fact by asking for any fact which matches a pattern expressed in an appropriate pattern matching language. (In particular the pattern can have variables in it so an entire fact can be retrieved by knowing part of it.)

Inferences about the story are also made on a line by line basis rather than waiting for a question to be asked. The inferences are also stored in the data base, and are expressed in the same semantic representation.

Since, in general there can be an infinite number of inferences from a given story, only the "important" ones are made at story reading time. If a

3These minimal requirements are met by virtually all specific representations proposed in the AI literature, as for example in (Rumelhart '72), (Sandewall '70), (Schank '72), (Simmons '70), and (Winograd '72). Representations which meet
emphasize this fact all rules will be written with single quotes around them. Single quotes should then be interpreted as saying, "consider all herein expressed in a suitable formalism".

Everything just said about the informal expression of rules applies equally well to the informal expression of semantic representation. So whenever it is necessary to talk about what appears in the data base, English augmented with parentheses and assertion names will be used, again encapsulated in single quotes.

Double quotes will have their normal panoply of uses plus one idiosyncratic use. Sometimes in order to refer to an entire sentence, only a double quote of a key word or phrase will be given. In such cases the sentence should be read as if the quoted key word (or phrase) were replaced with the original sentence. To make this slightly more complicated, sometimes a double quoted section will have single quotes around it, which means that the English sentence really represents a piece of formalism. For example, suppose we were talking about a story where we are told that Jack was angry, and Jack kicked Bill. We might say something like this: Naturally it is important to infer that "angry" was probably true before "kick". Furthermore we must make this inference even if "'kick'" appears in the data base before "'angry'".

\[5\text{Note that Nl is a variable which represents an assertion number, while nl would be a particular constant assertion number.}\]
§2 TWO PATHS TO RE-EXCHANGE

§2.1 The Role of "take back"

The most obvious hypothesis which would explain our understanding of "exchange at store" in (1) is that "take back" somehow means "exchange". Before trying to formulate that, however, one should distinguish between several different readings for "take back".

The reading one normally gets in (1) could be called the take to reading, since it can take a "to LOCATION" after it, as in "take back to the store". In this reading the PERSON has an OBJECT which he is transporting to a LOCATION.

Another common reading is the take from reading, as it can be followed with a "from PERSON2", as in "Jack took the ball back from Bill". In appropriate circumstances we can delete the "from PERSON2" and still retain the take from reading, as in:

(2) First Jack gave Bill the ball, then he took it back.

A reading similar to the take from reading is the retract reading, as in "take back the insult". It is similar in that in some sense one must "give" an insult before it can be taken back, but distinct in both its intuitive meaning and syntactic properties (the retract reading cannot take a "from PERSON2").

(3) *Jack took the insult back from George.

In all the above readings "back" acts like a particle. That is, "back" displays particle movement.

(4) Jack took back the insult.
(5) Jack took the insult back.

The final reading to distinguish is a non-particle reading.

(6) I took the children to Yellowstone Park last year
       and I will take the kids back this year.

(7) *I took the children to Yellowstone Park last year
       and I will take back the kids this year.

Although it is difficult to account for the ability to decide
which reading is intended in a given story, this problem will
not be considered here. Rather I will act as if the take to
reading were the only one possible in (1).

   Even after distinguishing between these readings, it is
still clear that "take back" cannot, in general, be replaced
by "exchange at store". Consider:

(8) Penny found a toad at the swamp. The next day her
       mother made her take it back.

(9) Jack was outside with his wagon. Bill and George
       wanted him to go to the circus with them. "Wait
       here while I take back my wagon" said Jack.

(10) Jack borrowed George's camera. The next day he
      took it back.

(11) Janet took her painting out of an exhibit at a local
      gallery because she decided that the painting was
      not good enough. However, an art critic's praise
      convinced her to take it back.

In fact, when using "back" it does not even seem necessary for
OBJECT to have been previously at LOCATION. Assuming Janet
lives in the U.S.A. we can have:

(12) When Janet heard what U.S. customs would do with it
she decided not take back the salami she bought in Italy.

The diversity of such examples might lead one to the following conclusion.

(13) 'PERSON take back OBJECT to LOCATION has no significance other than the trivial one of PERSON take OBJECT to LOCATION where PERSON or OBJECT has been before'. (Hence any other significance is due to context.)

That is, "take back" is the same as "take", except for a restriction on possible locations.

However, there is evidence against (13). Consider:

(14) Jack bought a top. The next day he took it back.

In (14) we have pared the information down to the bone. We are only told that the top was bought, yet we still understand "take back" as exchange. Note however what happens when we replace "take back" with "take to the store", as in:

(15) Jack bought a top. The next day he took it to the store.

In (15) we are left uninformed about why the top is being taken to the store.\(^\text{6}\) Or consider the following pair of stories:

(16) Janet bought a purse at the local store. The next

---

\(^{6}\)The lack of specificity introduced by replacing "back" might be even worse than (15) would indicate. For example it is not clear how we are able to figure out that LOCATION='store'. If we really believed that all "back" signified was a location where object had been before, then (14) would be more accurately paraphrased by:
day she decided she wanted a doll, so she picked up her new purse and took it to the store.

(17) Janet bought a purse at the local store. The next day she decided she wanted a doll, so she picked up her new purse and took it back.

In (17) we assume that the purse will be exchanged, but not in (16). Since the only difference is "back" replacing "to the store", somehow "back" must influence us so that we infer exchange. Another piece of evidence:

(18) Today was Jack's birthday. Janet and Penny went to the store. They were going to get presents. "I will get a top," said Janet. "Don't do that," said Penny. "Jack has a top. He will make you take it to the store."

Notice how strange the last line sounds when we replace "back" with "to the store". To explain the differences we see between (1) vs. (18), (14) vs. (15), and (16) vs. (17) we must assume that "take back" has more significance than hypothesis (13) allows. The question is, what significance?

Looking back at our earlier "take back" examples there

(15A) Jack bought a top. The next day he took it to a place it had been before.

Example (15A) is even less clear about Jack's motivation than (15).

7 There are individual differences in the interpretation of (16) and (17). Some find (16) uncertain between the exchange and non-exchange readings, but (17) they clearly interpret as exchange. Others see (16) as non-exchange, but (17) as
is one pattern which emerges. In those cases where LOCATION is where PERSON got OBJECT, the intent is to undo whatever he was doing when he got the object in the first place. So in taking the frog to the swamp, (11), Penny intends to free it--undo the original capture. In taking back the painting to the the exhibit, (14), Janet is re-exhibiting it.

Writing this up as an admittedly vague rule, we get:

(19) The Take Back Rule (Version I) 'PERSON take back OBJECT to LOCATION where he got it implies that he will undo what he was up to when he got OBJECT originally'

In §4 we will substantially alter this rule, but it will do for now.

Now in the case of taking back to the store, rule (19) is not enough by itself. Notice that when we exchange something at the store we not simply undoing the original "buy". For example, we don't get back the original bills and coins. In fact, often what we get back is not money at all, but rather a credit slip, or a different object. What this means is that for rule (19) to work for (1) we also need information associated

---

uncertain. For all people however (17) is more likely to mean exchange than (16). This conclusion is sufficient for the argument.

8 The English verbs used in the expression of (19) have tense associated with them so it might appear that (19) is only applicable when PERSON has already bought OBJECT but not yet exchanged it. Naturally this is not what is intended since in (1), for example, Janet has yet to buy the top. The rules then must be kept flexible with respect to tense
with buying things at stores which says that undoing a buy is a new activity, which we will call \textit{re-exchange}. That is, we need a rule like:

\begin{itemize}
\item[(20)] 'The undo of buy at store is re-exchange'.
\end{itemize}

Then, associated with 're-exchange' we have information about the kinds of items the buyer might get back, as well as any other facts we know about the activity. Intuitively then (any precision must wait until §4) rules (19) and (20) combine to interpret the last line of (1) as "'re-exchange'". (This path to 're-exchange' will be called the \textit{take back path}.) At the same time (19) is general enough to encompass the other examples of "take back". But as we are about to see, this is far from a complete explication of the last line of (1).

\textbf{§2.2 The Role of Context}

One problem with our explanation is that while (18) (our original story only with "take to store" replacing "take back") is distinctly odd, it is nevertheless understandable. That is, we still know that what Penny meant to say was that Jack would make Janet exchange the top at the store. Rule (19) cannot account for our understanding of (18) simply because it predicts that "take back" is crucial to our understanding, and (18) is missing "take back". Furthermore, we can't simply generalize rule (19) to include "take to", because then we would no longer (although accomplishing this will probably not be an easy task). However, to use English which reflects this timelessness would make the rules at best stilted, and at worst unintelligible, so no attempt will be made to do so.
be able to account for the differences between "take to" and "take back". To make things more complex there are even examples where "take to" is used in a situation like story (1), yet the example sounds much better than (18). Consider:

(21) Today was Jack's birthday. Penny and Janet went to the store. They were going to get presents. Janet decided to get a top. "Don't do that," said Penny. "Jack has a top. I know what will happen. He will thank you, and then the next morning, right after breakfast, he will take the top to the store."

While it is probably true that, when appropriate, "take back" always sounds better than "take to", examples like (21) seem to be within the bounds of acceptability.

Now one difference between (21) and (18) is that while in (18) it is Janet who will be going to the store, in (21) it is Jack. We can show that this is a not a major factor by noting the oddness of the following:

(22) Today was Jack's birthday. Penny and Janet went to the store. They were going to get presents. Janet decided to get a top. "Don't do that," said Penny. "Jack has a top. He will take it to the store."

As one might expect, the same phenomenon turns up in other "take back" stories. For example:

(23) Janet took her painting out of an exhibit at a local art gallery because she decided that the painting was not good enough. However, an art critic's praise convinced her to take it to the gallery.
(24) Janet took her painting out of an exhibit at a local art gallery because she decided that the painting was not good enough. However, an art critic's praise convinced her that it was really quite good. That afternoon, after eating lunch, she took the painting to the gallery.

Again, (23) is distinctly odd, but (24) is better. To explain these examples we need first to account for our ability to understand them at all, and then explain why padding the story with extra facts makes "take to" more acceptable than it is in the unpadded version.

Since, as has already been pointed out, we cannot generalize our "take back" rule without incurring trouble elsewhere, we need a new rule to explain the stories with "take to" (which will be called backless stories). One clue to the nature of this rule comes from an observation we made earlier. In example (15), reproduced here as (25), we found "take to" inconclusive about whether re-exchange is intended.

(25) Jack bought a top. The next day he took it to the store.

The main difference between (25) and (21) (a backless story where we do not understand that re-exchange is intended) is that (25) is missing the reason why the object is going to be re-exchanged. Indeed, if we add a reason why Jack would want to exchange the top to (25) we do get the exchange interpretation.

(26) Jack bought a top. When he got home he found it had a crack in it. The next day he took it to the store.
In exactly the same manner, if we take (24) (an acceptable back-less story) and remove the reason for taking back, we find the story becomes inconclusive.

(27) Janet took her painting our of an art exhibit at a local art gallery because she decided that the painting was not good enough. That afternoon, after eating lunch, she took the painting to the gallery.

We can codify these observations in the following rule.

(28) "take to" implies re-exchange in the presence of a reason for re-exchange'

However, we can significantly generalize on this rule.

Notice that taking the top to the store is really part of the act of re-exchanging it. (More precisely, arranging that the top gets to the store is part of re-exchanging it.)

This suggests that perhaps the real rule is of the form:

(29) 'A sub-action of A implies A when in the presence of a reason for A'

Rule (29) is supported by other instances where we infer what a person is really doing. For example:

(30) The play house had a loose board. Jack got a hammer.
(31) It was stuffy in the room. Jack went to the window.
(32) There was a sale at Macy's. Ms. Jones put on her coat.
(33) Jack's glass had stale beer in it. He walked over to the sink.

Formalizing rule (29) somewhat we get:

(34) The R+SSA rule

'If the story gives information which would make it plausible to infer that PERSON is favorably in-
clined toward action A, and PERSON does S, a significant sub-action (SSA) of A, then infer that PERSON is doing A'.

There are two important differences between (34) and (29). First, rather than talk about a reason for doing A, (34) just mentions information which would make PERSON wanting A reasonable. This allows for the possibility that there is no one sufficient reason for A, but several "demi-reasons" (more on this in §3.3 and §3.6). Second, the term significant sub-action is introduced because there are non-significant sub-actions which (34) should ignore. For example, if we know the play house is outside, it is the case that one has to be outside before one can fix the play house. Nevertheless, (34) does not apply in:

(35) There was a board loose in the play house. Jack went outside.

By itself, (34) is not enough to explain our backless stories. For one thing we need the rule:

(36) 'Getting OBJECT to STORE is a significant sub-action of re-exchanging OBJECT at STORE.'

Naturally this rule will only work for re-exchange stories. However (36) (or a generalization of it, if one exists) is needed independently of our troubles with (21) (backless version of (1)) as can be seen in:

(37) Jack was going to exchange his top. He took it to the store.

Rule (36) explains why he took it to the store.⁹

⁹The evidence for rules (34) and (36) seems quite strong,
Nor is (36) sufficient to allow the use of R+SSA, since we also need to infer that the person who did the sub-action of A might want A itself. If we are told that PERSON wants A this is easy, but this is not what happens in (21) and (24), the backless stories we wish to explain. Instead we must construct a plausible scenario leading from the known facts to PERSON wanting A. In an example like (24) the general logic is clear:

(38) 'Janet's negative opinion of the painting is the reason for her desire that the painting not be in the show. Janet changes her opinion of the painting, so it is reasonable that she now wants the painting in the show. One way to accomplish this is to take it there and re-enter it.'

As we will see in §3.6, (38) hides many complexities, and the in that there is evidence independent of story (1) that each is needed. There is however one problem. Consider the following example:

(37A) Jack got a rock in the woods. However, his mother would not let him keep it. That afternoon he took the rock to the place in the woods where found it. While superficially this is somewhat similar to the other "take to prior location" stories, rule (34) does not seem to apply, or at least not very easily. The problem is that taking a rock to the place where it was found is not normally part of getting rid of a rock. Nevertheless, we understand Jack's activity as part of rock abandonment, although we wonder why he is going to all that trouble.

One possible way out is to view (37A) as evidence for a new rule to replace (36) (the rule which states that "taking" is part of 're-exchange').

20
problems in working out the scenario for (21) is even more dif-
ficult. First, we find that to model the reasoning in (21) 
after that in (38) we need the statement 'Janet's belief that 
Jack wants the top is the reason that she wants it.' Nothing 
like this appears in (21) of course, so this must be another 
piece of knowledge available to the model. Formulating this 
rule will be a major task in §3. Secondly, it turns out that

(37B) 'If OBJ was obtained at LOC as part of activity A, 
then taking OBJ to LOC is an SSA of undoing 'A' 
The trouble with this rule is that it makes no distinction 
between cases like "taking the rock" where we don't understand 
why the person is going to so much trouble, from "taking in 
order to exchange" where we know that the taking is a require-
ment, to "taking a frog to the swamp" where the taking is not 
required, but is considered a nice thing to do, certainly from 
the frog's viewpoint. To account for these differences seems 
to require separate rules like (36). If this is the case, then 
(37B) would be essentially an ad hoc rule, only used to explain 
(37A). Furthermore, while criticizing the take back rule in 
§4 we will show that the relation between 'taking OBJ to LOC' 
and 'undoing A' is problematic. These problems apply equally 
well to (37B), so (36) will be retained instead. As for (37A), 
the only explanation I can envision at this point is that we have 
a rule of putting things back where they came from, and somehow 
this rule allows R+SSA to apply to (37A).

10 Actually, we also need 'Jack does not want a top' and in 
(1) we are not even told that. Rather we are told that he al-
ready has one. However it seems reasonable to assume that 
Jack's not wanting another top is deduced when Penny says that 
he already has one. The problem then reduces to our not being 
told that Janet is getting the top because Jack will like it.
there is more than one plausible scenario leading from the known facts to Janet's taking the top back. Describing both scenarios and the interaction between them will also be done in §3. In the meantime, to get on with the current task of showing how the R+SSA rule gives a second path to "re-exchange" we will propose the following dummy rule:

(39) This rule, or collection of rules, will allow us to interpret "Jack does not want a top" as grounds for Janet's re-exchanging the top in stories like (21).

With this dummy rule, we can now show the second path leading to the conclusion that the top will be exchanged at the store.

(40) 'Jack does not want a top and rule (39) imply n1(Janet has reason to want to re-exchange the top). Janet will take the top to the store and rule (36) imply n2 (Janet will perform an SSA of Janet re-exchanging the top). n1 and n2 and the R+SSA rule imply Janet will re-exchange the top at the store.'

The path to 're-exchange' described by (40) will be called the R+SSA path.

§2.3 Two Conclusions

The two paths to 're-exchange', "take back" and R+SSA, constitute the redundancy mentioned in the introduction. Of course, there were some stories which only had one operational path. Indeed, it is hard to imagine being able to distinguish the two paths if this were not the case. However, given the existence of single path stories it might be claimed that all we have
found is a story (1) with too much information—like a mystery story with too many clues. This, naturally, is a possible interpretation of the example. I do not choose this interpretation myself because it seems to me that is story (1) which is the natural case, and stories like (14) and (26) which seem pared down and unnatural. To put this another way, were we to ask people to construct stories about exchanging things at stores, I feel confident that (1), rather than (14) or (26) would serve as the prototype. When we find then that all mysteries have exactly the same extra clues, we are entitled to conclude that we are observing not a property of the particular stories, but rather of their creators.

Of course, semantic-pragmatic redundancy is not a new phenomenon. Consider a typical case:

(41) Jack is young but has grey hair.

In (41) the "but" tends to reinforce our conviction that it is somewhat unexpected for a young person to have grey hair. Since we would have deduced this oddness even without it, the "but" is redundant. However, here "but" merely reinforces. We could not expect to include the "but" while leaving out the last clause and have the reader figure out that Jack has grey hair.\(^\text{11}\) In (1) either set of facts are sufficient to

\(^{11}\) This is not say that "but" always plays a reinforcing role. While I know of no example, I would not be surprised to find cases where the addition of "but" leads to a different inference. My point is rather that "but" usually plays a reinforcing role because the amount of semantic information it
make the inference. What we observe in (1) then is an extreme case of semantic-pragmatic redundancy. It is of interest because, to my knowledge, it is the most extreme case yet observed.

As yet we still have not completely explained example (18). In this example we saw that if Penny replaced "take it back" with "take it to the store" the story sounded strange. One might even say that Penny's utterance was ungrammatical. It cannot simply be some grammatical constraint of the form "you must use 'back' if OBJECT has been at LOCATION before", or "you must use 'back' if what is going to happen is an exchange". Such rules would predict, incorrectly, that (21) and (24) (backless stories with filler) would sound as odd as (18) and (23).

Looking just at (18) and (21), the real difference seems to be that in (18) Penny simply tells Janet why Janet should not buy a top, while in (21) Penny tells Janet a story, the moral of which is that Janet should not buy a top. That is, in (18) Penny is in conversational mode, while in (21) she is in narrative mode. On the other hand, notice that by removing "back" we are also removing one path to the conclusion that the top will be exchanged. Now all inferences have some degree of doubt attached to them. Presumably, the certainty contains is small. However even a small amount of information may, in some cases, be enough to push through an otherwise shaky inference.
we feel that the last line of (1) implies "re-exchange" is due to our having two independent ways of coming to that conclusion. By removing one of these paths we reduce our certainty, as can be observed in (14) and (18). (The first has "back" but no reason for the exchange, while the other has the reverse.) These observations lead to the following hypothesis as a means of explaining the observed differences between (18) and (21):

(42) The level of "surety" for inferences must be higher in conversational mode than in narrative mode to keep a constant degree of naturalness.

To put this slightly differently, inferences we would make in the course of a narration will sound strained if required in the midst of a conversation.

Naturally this rule correctly predicts that (1) and (21) will both sound better than (18). However, there is further evidence which supports it. Consider our original story (1) but with the last line now changed to:

(43) Jack will make you take it to the store and get something else.

(44) Jack will make you take it to the store and exchange it.

These lines sound perfectly reasonable in a conversational context. Rule (42) predicts this since in both (43) and (44) we are given further information which supports the hypothesis that re-exchange is intended. In (44) aside from the ambiguity of "exchange," we have spelled it out. In (43) we need an
inference to go from "get something else" to "re-exchange", but it is an inference we need to make anyhow (else we might think that Jack was asking for two gifts).

Further evidence comes from other cases, (30)-(33), where we make an inference using our R+SSA rule. Most of these examples, when modified so that the deduction occurs in a conversational mode, sound odd. Consider:

(45) It was stuffy in the room. Mother suggested that Jack go to a window.
(46) There was a sale at Macy's. Mr. Jones suggested that Ms. Jones put on her coat.
(47) Jack's glass had stale beer in it. His girlfriend suggested that he walk over to the sink.

Unfortunately, not all our R+SSA examples have this property.

(48) The play house had a loose board in it. Mother suggested that Jack get a hammer.

While this has a slight odd sound to my ears, it is clearly not as bad as (45)-(47). Nor is it a unique case.

(49) Jack needed money. Mother suggested that he get his piggy bank.

In both (48) and (49) we have a person getting an object which is only used for certain purposes. Hence in both stories there is a close fit between the reason and what the object is good for. This would suggest that the inferences in (48) and (49) are simply more certain than (45)-(47) and hence the stories sound better. This is supported by:

(50) Bill called Jack asking for the return of the ten books he loaned him. Jack went looking for a large box.
(51) Bill called Jack asking for the return of the ten books he loaned him. Bill suggested that Jack go looking for a large box.

That (51) sounds much worse than (48), (49), and (50) is presumably due to the fact that boxes are not used for returning books in the same strong sense that hammers are used for fixing loose boards, or piggy banks used for obtaining money.\footnote{Let me use this footnote to mention two theories which have been suggested to me, but which do not account for the data. The first is that a rule on the order of "you must say all you know" as suggested by Grice (Grice '68) would predict the badness of (18). While this is true, it would also predict that (21) should sound just as bad which is not the case. We could patch this with an addition to the rule which says that it can be ignored to some extent while in narrative mode, but this would still fail to explain why (43) and (44) are good. The second suggestion was that the repetition of "the store" in (21) violates rules of pronominalization, and hence the story}}
rule. The question is, why does the story sound worse with "take back" missing than with R+SSA missing? One possible explanation is that the R+SSA path is less sure than "take back", but barring further evidence, this is at best ad hoc. A more likely possibility is that Grice's rule of "saying all you know" (Grice '68) is applicable here. So, by using "take to" rather than "take back" one is telling the listener of the sub-action rather than the real action that is important. If it is correct to use Grice's rule in this manner it would be natural for (18) to sound worse than (52) since both have roughly the same level of surety, but (18) also breaks Grice's rule. However, this all seems quite speculative at this point.

sounds bad. While such repetition can make a story sound odd, that is not the cause here since we could have "Jack will exchange it at the store" and the story would sound fine. To clinch the case, note the following two paradigm cases.

(51A) Janet said to Penny, "I got this top for Jack for his birthday." "Where did you get it," said Penny. "At Joe's Toy Store," said Janet. "Jack doesn't want a top," said Penny. "He will make you take it back there *to the store back to the store

(51B) Jack traded his pocket knife to Bill for Bill's top. When Jack discovered that the top had a crack in it he took it *to Bill back to Bill back to Bill
The second problem with (42) is precisely distinguishing narrative from conversational discourse. So far I have relied on the reader's intuition as to what these terms should subsume. Some obvious factors are:

(53) One person telling another something as indicated by quotation.
(54) One person telling another something as indicated by the scope of a verb indicating speech.
(55) Style.

(53) is straightforward except for the fact that (55) can override it. (54) is also reasonable enough as far as it goes, but it may not go far enough. Although at this point I no longer trust my intuitions on these examples, I have a suspicion that phrases within the scope of other verbs also are in conversational mode, for example, "decide" as in "Jack decided to take the top to the store." Whether this true or not I leave to people with less biased ears than mine.\textsuperscript{13} Finally, (55) is

\textsuperscript{13}Since making the observations found in this section I have come across (Kuno '72) a paper which has direct relevance to the analysis of conversational modes presented here. Firstly, Kuno isolates a class of verbs which he claims are best represented (in deep structure) in a "direct discourse representation". For example, (55A) would be represented as something like (55B).

(55A) John expects to be elected.
(55B) John expects, "I will be elected."

(Kuno points out that there are several possible ways to represent the basic idea of (55B). Neither his, nor my observations depend on, for example, the presence of quote marks.) Naturally, "decide" would be one of these verbs, and it may well turn out
clearly needed but very vague. As was mentioned when we first encountered acceptable backless stories, even though Penny was still talking to Janet (as indicated by quote marks) the presence of extra facts made her comment into a story within a story. How this sense of style is to be formalized I do not know.

that the class of verbs which put one in conversational mode are exactly Kuno's direct discourse verbs. Secondly, Kuno mentions a paper, (Kuroda '71) which discusses two styles of story telling in Japanese, "reportive and nonreportive", and finds syntactic differences which correlate with this distinction. I have not read (Kuroda '71) myself, but from Kuno's description "reportive" seems similar to what I call "conversational", and "nonreportive", "narrative". Naturally, if this were true it would be of interest.
§3 HOW WE DECIDE THAT JANET HAS A REASON FOR TAKING THE TOP BACK

§3.1 Who is Taking it Back

In §2 we noted that our theory required us to deduce the likelihood of Janet's wanting to take the top back. This was because the R+SSA rule (which took us from "take to store" to "re-exchange") required added evidence that this was indeed what was going on. We then commented that for the stories we were concerned with, inferring likelihood amounted to producing a plausible scenario starting from the facts in the story and leading to the act in question. However we did not construct such a scenario for (1) in §2, and this is the task we will now attack.

To give ourselves a starting point let us state the R+SSA rule slightly more succinctly than we did in (34).

(56) 'If it is plausible that Pl positively evaluates action A and Pl does A1, an SSA of A then Pl does A'  

To this let us add a second rule which will give in some circumstances a reason for taking back.

(57) 'If PERSON negatively evaluates object X then PERSON might positively evaluate re-exchanging X'  

(Note that we have introduced a new term here, positively/negatively evaluate. While we will be somewhat more precise about its definition later on, for the moment it will suffice to say that it is intended as a general purpose term to cover ideas like "want/not want".) Now (57) will work in examples where Jack does not like the top and proceeds to take it back himself. However, note that if it is Jack who does not like his
top, but Janet who takes it back, (57) will not provide the
necessary inference for (56) since it would give a reason for
Jack's wanting it done, whereas (56) needs a reason for Janet's
wanting it done.

This is, of course, due to the fact that (56) states that
the person with the reason must be the actor, and (57) states
that the person who does not want the object is the person
who wants the re-exchange. Note, however, that if we were
simply to loosen either of these conditions the model could
then handle stories like (1) where Janet does the taking
back. That is, it could handle (1) if we adopted either

(58) 'If it is plausible that P1 positively evaluates
action A and P2 does A1, an SSA of A then P2 does A
or

(59) 'If P3 negatively evaluates object X then P4 might
positively evaluate re-exchanging X'

We only need adopt one of the two since if, for example, we
adopted (59) it would allow the model to have scenarios in which
Jack's not liking the top is reason for Janet's taking it back,
which would then satisfy the strict R+SSA rule. The same would
be true if we adopted (58) and (57).

But accounting for (1) by simply adopting (58) and (59)
would be a mistake. Consider what happens when it is neither
Janet nor Jack who takes the top back. Say it is some person
Frank, who has been introduced into the story earlier, but
who has nothing to do with Janet's giving Jack the top. It
is true that people can understand such stories, but only by
assuming that Jack asked Frank to do it. Furthermore, every-
one agrees that such stories are much less natural. Finally, if we had a story where Frank did not know Jack at all, then we could not assume that Jack asked him to do it, so we would be left with no explanation at all, and a very odd story. The point is that if we simply adopt (58) or (59) we cannot account for any of these facts. By adopting, say, (58) we are in effect saying that anybody can take the top back, and that we recognize no distinction. That this is not the case argues against loosening the restrictions on either (56) or (57). Furthermore, the fact that people assume that Jack asked Frank to take the top back is quite significant. What it seems to imply is that we were looking for a reason for Frank's taking the top to the store. We could not find one, but we noted that Jack had a reason to do it, so we invoked a rule something like:

(60) 'To infer P1 positively evaluating doing X, if there is a reason for P2 doing so, then assume P2 asked P1 to do it for him'

We have the same situation with other cases where we use the R+SSA rule. For example:

(61) Jack thought it was stuffy in the room. He walked over to the window.

(62) Mother thought it was stuffy in the room. Jack walked over to the window.

If we are to understand (62) at all we must assume that Mother asked Jack to do it. This is easily accounted for in the model if we assume the strict forms of (56) and (57) (or rather the equivalent to (57) for stuffy rooms) and the use of (60) to overcome the strict requirements on who it is that does the wanting.
So far the discussion has tacitly assumed that the only two options were either keeping (56)-(57) intact, or making a single change in either adopting (58) or (59). In fact there are other possibilities. Suppose what Penny had said to Janet was:

(63) "You will end up taking the top back because Jack has a top."

If our system is to understand this statement it should not take the asserted causal relationship too literally. Instead it should realize that there are several missing links in the causal chain leading from Jack having a top to Janet taking the top back to the store. But to recognize this fact we need rules for defining "allowable causal relations". (See (Schank '73) for a more detailed discussion of this issue.) Given the need for these rules anyway, we could allow either (58) or (59), but the causal relation they posit will be checked by the "allowable causal checker", which, seeing that they are not allowable, will try to fill in the missing steps in the causal chain.

However, from the viewpoint of what we are trying to do in this section it really doesn't make much difference whether we adopt this solution, or the original one of keeping (56) and (57). In both cases the model will need rules which will explain why it is that we consider it reasonable for Janet to take the top back. So let us return to this task.

Since in (1) we do understand that Jack asked Janet to re-exchange the top, this might indicate that (60) is used when inferring why Janet would take the top back. But (60) by it-
self is not sufficient to explain all our intuitions. We noted earlier that Frank's taking the top back was much less natural than Janet's doing it. But (60) would still not predict this, since as far as it is concerned it is as reasonable for Jack to ask Frank as Janet (assuming, say, that he knows them both equally well). But people do not see Janet and Frank as equivalent in this situation. There is some logic in Janet's doing the deed which simply is not present in the case of Frank. So there must be further rules which make the scenario leading to Janet's taking the top to the store more plausible than Frank's doing it. If you ask people why there should be this difference, they will usually say something like "If you get something for somebody you want them to like it".

Formalizing this common sense slightly we get:

\[(64) \quad \text{'If PERSON gets OBJ in order to give it to PERSON2 and PERSON2}^{+} \text{ OBJ then PERSON}^{-} \text{ OBJ.}^{14}\]

---

\[^{14}\text{In this rule it is necessary to assume that PERSON knows that PERSON2 does or does not want OBJ, because if PERSON does not know then his opinion of OBJ would obviously not be affected. This problem of assuming knowledge on the part of the characters in the story, like the problem of tense, appears in many ways in many different rules, and, as in the problem of tense it will be assumed the solution will not greatly affect any particular rule, hence making it safe to ignore the problem in this paper.}\]
(I have here introduced the abbreviation + for positively evaluate and - for negatively evaluate. I am also assuming a notation for alternatives in rules so if + is chosen on the first occasion it must also be chosen on the second. The same is true for -.) With the addition of (64) the model can infer that Janet might now negatively evaluate the top and hence, by (57) she might exchange it at the store.

Although both (57) and (64) will be substantially changed before we are done (neither captures all the right generalizations), it seems clear that we are moving in the right direction. In particular we are at least now in a position to account for the fact that while we can understand Jack, Janet or Frank taking the top back, Jack seems more reasonable than Janet, who in turn is more reasonable than Frank. (Actually, to do this the model also needs a rule which says that if a person wants something to happen, everything else being equal, he is the most likely person to do it.)

Before we go on to improve our rules, I should point out that many of the rules to be developed in this section are to some extent idiosyncratic, and at the very least culture dependent. So, for example, in the culture I come from, which is what will be modeled in this section, it is the case that Jack's taking the top back would be more common than Janet's doing so. However, I have talked with some people (Swiss, although they claim that their impressions are typical for Europeans) for whom in example (1) it would be more common to have Janet take the top back. They explain that the store owner would recognize Janet and hence be willing to exchange
the item, but he would be reluctant to do the same for Jack. As mentioned in the introduction the basic model must be flexible enough to accommodate such differences. In this case the difference would probably manifest itself in other rules concerning how to re-exchange something. In my culture there would be checks on how recently the object was purchased, whether the owner was responsible for any damage to the object and perhaps others. The European version of these conditions would simply also include the expectation that the person who does the re-exchanging is the person who bought the object at the store. This rule would then take precedence over the assumption that the person who wants the action (the re-exchange) is the most likely to do it.

Prior to this section the major rules under discussion have not been so culturally dependent. The take back rule is a fact about the English use of the word "back", so is language dependent. It might prove to be culturally dependent also, but this is not obvious. The R+SSA rule on the other hand, is probably culture independent. There is nothing in the R+SSA rule which refers to objects or situations found in any particular culture. That such a rule could still be culture specific would be quite strange.

§3.2 Negative and Positive Evaluation

We now have two rules. Rule (64) makes the inference from 'Jack negatively evaluating the top' to 'Janet negatively evaluating the top' and (57) then allows the model to infer
'Janet positively evaluating re-exchanging the top'. Almost all the rest of §3 will be concerned with improving these rules. This section and the next will deal with (57), after which we will take up (64).

One minor change which is needed is in the use of negative and positive evaluation. In (56) and (57) the terms are applied both to objects and to actions. Instead of allowing the terms to refer directly to an object we will have it refer to states or actions which mention the object. By doing this we can also rid the model of one predicate by defining negative evaluation of a state or action X as positive evaluation of 'not X'. In particular, "Jack does not want a top" would cause the model to assert that he positively evaluates not having another top. (Actually things are probably more complicated. My current guess is that it would be represented as Jack positively evaluating not getting a top, and we would infer that the reason is that he positively evaluates not having the top which he would get.)*

---

*This is due to a conversation with Chris Riesbeck. The reason for doing things in this slightly convoluted way is as follows. First, intuitively when we say Jack does not want a top, we seem to be talking about getting one. Second, and more formally, if the model is then told that Jack has a top we do not want it to infer that he does not want that one either. If "not want a top" were translated into "not want to have a top" then the model would make this incorrect inference. (Naturally, if the phrase became 'not want to get a top' the "get" is in the future, and the "get" responsible for the top Jack has must be in the past, hence the top he
While this change seems intuitively reasonable, and many others have adopted it before me, let me just point out that something like it is needed if we are to represent the ambiguity demonstrated in the next example.

(65) "I have to buy Jack a present. Would he like a film to add to his old film collection? asked Janet.

"Jack doesn't want a Chaplin film," said Penny.

"What movie shall we see tonight?" asked Janet.

The line on the right can be interpreted two ways, depending on which of the fragments on the left precedes it.

Restricting the use of positive evaluation this way makes it easier to define the term more precisely. The idea we want to capture is that a person positively evaluates a state or an action (or the negation of a state or action) X, if he prefers X to not X. His preference for X can be quite weak however. So, if Jack intends to do X, but really does not want to, we

has does not qualify as an instance of the kind of top he does not want.) Furthermore, if we only learn that Jack has a top after we are told that he does not want one, then the model cannot translate "not want a top" into "not want another top" since it will not have yet learned of the one he has. This is the justification for translating "not want a top" as "not want to get a top". On the other hand, we cannot simply stop
will still say that he positively evaluates X. Hence:

(66) 'Pl intend Y implies Pl + Y'

(67) 'Pl want Y implies Pl + Y unless Pl not intend Y'

Finally, by representing \(+\) OBJECT as \(+\) some action or state mentioning the object, we can now generalize on rule (57) (which connects 'not wanting X' with 're-exchanging X') as follows:

(68) 'Re-exchanging OBJ has a result that the owner prior to the act no longer owns OBJ'

(69) If \(X\) is a state which results from action \(A\), then \(P\{\(+\}\) \(X\) is a potential reason for \(P\{\(+\}\) \(A\).

The idea behind (69) is quite simple. In the case at hand, for example, the model can conclude that Jack's not wanting the top is a reason for him to exchange it, assuming that he is the top's owner. More generally, (69) expresses our belief that people do actions to bring about desired states of affairs.

---

here. The model still needs to infer 'Jack not want to have a top'. The reason is that it must be able to understand what happens once Jack gets the top. For contrast, suppose we were told that Jack does not want to take a walk. If Jack is then forced to do so, we might conclude that he did not enjoy it, but we would not be on the lookout for further actions aimed at somehow undoing the result of taking a walk. After all, the walk is over, and while Jack might not have liked it, there is nothing he can do now. In the same way, if we leave "Jack does not want a top" translated only as "not get a top", once he has gotten it the act is over and there is nothing he can do about it. Of course, we all know that this is really not quite true. Jack can try to get rid of the top. But this does not solve the problem of getting it, but rather of having it, hence
§3.3 Evidence and Motivation/Action Theory--Dodging Two Problems

While rule (69) is intuitively reasonable, it is hard to know exactly how to make use of it. One problem it raises is that of evidence.

Consider the situation where we know that a person did some action, but we do not know why. Often rule (69) will provide the answer. For example, skiing is normally done for its own sake, hence rule (69) does not apply. Yet, with proper evidence, say Janet starting to ski down the hill after she sees another skier collide with a tree, we will decide that (69) is applicable after all. Or again, we would normally assume that a person painted a chair in order to make it look nicer, one possible result. But if we are told that a person who often plays pranks painted the teacher's chair just before the teacher walked in, we would assume it was the resulting stickiness of the chair which was of interest. Again, in applying (69) we must be able to weigh the evidence presented in the story.

Such problems of weighing evidence are probably best formulated independently of particular rules like (69) since ultimately the need to infer that Jack does not want to have the object. Note that if it were the case that this deduction were false (maybe Jack simply finds getting presents unpleasant), then not wanting a top would not be a reason for re-exchanging it. Finally, since the top Jack does not want to have is the same one he does not want to get the model still will not make the false inference that he does not want the top he already has.
the evidence rules must apply not only to inferring motivation from action, but also inferring that an action took place from its results, or that the world was in a particular state because because we know of corollary facts. The R+SSA rule, for example, will ultimately become part of our common sense theory of evidence.

The creation of such a theory is, of course, a major task in itself, and will not be attempted here. To get around the lack of this theory, we will accompany rule (69) with two other rules which specify, for a few cases, how the more general rule should be used. These subsidiary rules are expected to eventually be subsumed under the more general theory of evidence.¹⁶

\begin{align}
(70) \quad & '\text{If } P \{+\} \text{ action } A, \text{ and } P \{+\} R, \text{ a result of } A \\
& \text{then assert that the latter is a reason for the former}\'
\end{align}

\begin{align}
(71) \quad & '\text{If } P + A \text{ (some action) and one of } A's \text{ results } R \\
& \text{is singled out as the usual reason for doing } A, \\
& \text{then barring explicit evidence to the contrary, or } \\
& \text{a known reason for doing } A, \text{ assert that } P + R \text{ is a } \\
& \text{reason for } P + A' \text{'}
\end{align}

¹⁶Originally the theory presented in this paper was to be concerned with both what the model needs to know and how it applies this information. By placing a theory of evidence between the knowledge and the actions of the model I have, in effect, moved the current research away from the second problem, how the information is used in the model. Although it was not my original intent, I am following McCarthy's lead in putting the epistemological problem before the heuristic problem (his terms) (McCarthy and Hayes '69).
So far we have been discussing the situation where we know what was done, but not necessarily why. The other side of (69) is where the model is given some motivation, and it must infer what action, if any, will be taken to satisfy it. One problem here is that it is not clear to what extent people anticipate such things. For example, if one is at a friend's house and he says that he is going to have a cup of coffee, intuitively one anticipates his going into the kitchen and making some. On the other hand, if one character in a story tells another that he needs a green chair, given no other information it is by no means obvious that one anticipates how he intends to obtain it. But the question of how and when such actions are anticipated is not crucial here as it does not play much of a role in the rest of the paper. To get around this question let us instead consider cases where the potential action has already been singled out. For example:

(72) Fred asked John, "Do you know if Ann intends to drive to New York soon." "Well, she would like to be in New York to see the new exhibit at the Museum of Modern Art," said John.

Fred's question sets up a possible action by Ann without commenting on whether or not it will take place. John's answer, tells us that Ann would like to be in New York, yet, we still do not know if Ann intends to go. Hence we cannot have any rule which simply allows the model to infer \( P + A \) from \( P + \) result of \( A \). Yet, the model must at the same time recognize that John's answer is not a non sequitur. It does indeed
bear on the question which was asked. We can represent this situation as follows.

(73) \( (\text{Ann} + \text{HYP (Ann drive to New York)}) \)

reason

\( (\text{Ann} + (\text{Ann be in New York})) \)

The HYP here stands for hypothetical, and in cases like (73) we will say that Ann has a hypothetical (or potential) reason for going to New York. Naturally, if we later learn that she has gone, the model will use some rule like:

(74) 'Potential reasons become actual reasons if the action is done'

(The reader should note that the term potential reason was used in the statement of (69). While at that point it was simply a descriptive term, it has now become, at least to some degree, a technical term. However, the meaning of (69) has not changed any in the process.)

Hypothetical reasons allow us to separate the issue of what is or is not a reason for doing something, (or wanting to do something) from the issue of what we consider sufficient reason to get us to do it (or to get us to want to do it). This latter question is, like the question of evidence, a tough one, and will not be discussed in the paper except in a peripheral way. And, like the theory of evidence, this problem will also be given its own title, the motivation/action theory.

Both of the problems here have been discussed in the context of rule (69), but they are not specifically problems with that rule. Not only will these same problems come up with later
rules, but they also occur with (57) the less general precursor of (69). Because of (57)'s limited capabilities the problems were simply less obvious.

§3.4 Why Janet Should Not Give Jack the Top

Briefly recapitulating, of the two rules we are now concerned with, (64) and (57), we have been working on (57) for the last two sections. In particular, by replacing 'not wanting an object' with 'not wanting to have (own) an object' we were able to replace (57) with a more general rule, (69) which associates wanting an action with wanting its results. This rule, however, is difficult to apply in practice (as we noted in §3.3) unless we have both a theory of evidence and an action/motivation theory.

Now let us turn to the other rule, (64) repeated here as (75).

(75) 'If PERSON gets OBJ in order to give it to PERSON2 and PERSON2 \(\{^+\}\) OBJ then PERSON \(\{^+\}\) OBJ'

In light of the new definition of negative evaluation, \(\{^+\}\) OBJ should be interpreted at \(\{^+\}\) (PERSON have OBJ). The first necessary change to this rule is in its last line.

(76) 'If PERSON gets OBJ in order to give it to PERSON2 and PERSON2 \(\{^+_\}\) OBJ then PERSON is \(\{^+_\}\) about giving OBJ to PERSON2'

To see that this formulation is preferable, consider the following stories.

(77) Janet bought a top for Jack for his birthday. When she learned that Jack did not want a top she thought,
"I don't want this top then."

(78) Janet was thinking of buying a top for Jack for his birthday. When she learned that he did not want a top the decided not to get it.

(79) Janet bought a top for Jack for his birthday. When she learned that he did not want a top she was unhappy.

(80) Janet bought a top for Jack for his birthday. When she learned that he did not want a top she decided that she would not give it to him.

While it is unlikely that a single rule could handle all these various situations, it should be intuitively obvious that a rule something like (75) lurks in each of these stories. In the case of (77) it applies more or less directly. That is, (75) accounts for Janet's unhappiness with the top, and that is what we observe in (77). In (78) we need an intervening rule, because (75) talks about negative evaluation of having the top, whereas in (78) Janet comments that she negatively evaluates the act of getting it. However, we already have the necessary rule in (69) which says, in essence, that if the model sees approval (disapproval) of both an action and its result then the approval (disapproval) of the result is the reason for the approval (disapproval) of the action. In example (78) the model also needs a rule which relates negatively evaluating getting the top to not intending to get the top, and again we already have the needed rule, this time in (66). Similarly, a simple rule about negative evaluation of a current or expected state leading to unhappiness will account for (79).

This leaves (80) which, as far as I can see, cannot be
accounted for by using (75). The problem is that the model has to derive "not give the top to Jack" from "not want the top". But this is a very strange deduction. In fact, in general the opposite will be the case; if you don't like an object you have, you are more likely to give it away, not less so. On the other hand, if we adopt the revision proposed in (76), example (80) falls out immediately. Of course, we hardly ever get anything for free, and by adopting (76) we must provide a rule which will connect "not want to give the top to Jack" with "not want to have the top". But this rule, at least on the surface, is reasonable. After all, the reason why Janet got the top in the first place was to give it away. Now that the reason for having it has gone away, it is reasonable that Janet no longer wants to have the top. On this basis (76) is adopted over (75).

§3.5 "Not Want" as a Potential Reason

While in §3.6 we will return to the task of formalizing the rule which will bridge the gap between "not give" and "not have", it will be easier if we first make some minor changes in rule (76).

Consider our state of knowledge when we have just learned that Jack does not want a top, as in:

(81) After Janet bought a top for Jack's birthday she learned that he did not want a top.

Suppose we now ask, "Will Janet give the top to Jack?" The best answer is "I don't know" although "No" is also acceptable. However, depending on how we formulated it, (76) would force
the model to answer either yes or no. "I don't know" would not be a possibility. If we formulated (76) so that it only asserted a causal connection between "not want" and "not give" after seeing both stated in the story then (76) would have the model answering "Yes" to the question. On the other hand, if (76) immediately asserts "Janet not give" after seeing "Jack not want", then the model answer "No".

The solution to this problem is to rewrite (76) so that instead of asserting "not giving" it proposes a hypothetical reason for not giving. That is, (76) becomes:

(82) 'If PERSON + N1 (PERSON give OBJ to PERSON2) and N2 (PERSON2 \{+\} OBJ) then assert N2 is a reason for (PERSON HYP\{+\} N1)'

By changing (76) so that it only asserts that there is a reason why PERSON might not want to give OBJ to PERSON2 we remove from the rule entirely the question of whether or not Janet will give Jack a top she knows he does not want. As we commented in §3.3, deciding on a character's course of action given only motivation and surrounding circumstances is a hard problem and one which is presumable independent of particular rules. Rule (82) will eventually need reformulation since an individual motivational rule such as (82) must tell the motivation/action theory how important it is that it, (82), be satisfied. But how such information should be formulated is a question which will have to await further knowledge about the nature of the motivation/action rules themselves.

Of course, passing the buck to the yet unformulated motivation/action theory is an unsatisfying thing to do. But it is
hard to see what other choice we have. Consider:

(83) Janet was considering buying a top for Jack's birthday when she learned that he did not want a top.

(84) Janet had already bought a top for Jack's birthday when she learned that he did not want a top.

(85) Penny and Janet were already a few minutes late to Jack's birthday party. On the way Penny asked Janet what she had bought Jack. When Janet said that it was a top Penny told her that Jack did not want a top.

While in (83) we feel certain that Janet will not give a top, (84) could be answered either way, and in (85) I would lean toward the conclusion that she will give Jack the top anyway, though perhaps along with an apology. The difference of course is the ease with which Janet can change her plans, or to put it another way, how much motivation Janet has to give Jack the top anyway. It seems clear that a single rule like (76) cannot be expected to handle such diversity, and a more complex set of motivation/action rules must be postulated.

What this change means in terms of the discussion to follow is that for an example like (84) we cannot assume that the model has inferred that Janet will not give the top to Jack. So if the next line of the story were "Janet took the top to the store", the search for a reason for her action becomes that much more difficult. For example, in the next section we will be concerned with inferring "Janet - top" from "Janet - give". We now see that we cannot even be sure of the latter.

On the other hand, the R+SSA rule as formulated in (34) only
requires a plausible scenario leading to 'PERSON + A' (in this case "Janet + re-exchange"). Since the reason the model is unsure about Janet's giving the top or not is that it is plausible either way, there is no reason in principle why '"Janet + give"' being *biplausible* should prevent it from satisfying the R+SSA rule (provided the model has the necessary inference mechanisms).

So, to model those people who answer "I don't know" to (84) we will assume that '"Janet + give"' is not marked false, but rather *biplausible*. Furthermore we will assume that our rules of inference are capable of inferring new biplausible assertions from old (one rule which must be able to do this is (91)), and that the R+SSA rule allows biplausible assertions, or their negations, to figure in plausible scenario construction.17

Given these assumptions it makes no difference that we do not know for certain that '"Janet - give"' and for ease of presentation the further development of our rules will take no cognizance of this fact.

17This issue is another part of evidence theory. One possible way to handle the problem is to create alternate possible worlds, one where Janet intends to give the top, another where she does not, so that within each world the model can act as if it is sure of its beliefs. For a model which works this way, plus one of the best discussion of problem of evidence in common sense systems, see (McDermot '73).
§3.6 How to Infer "re-exchange" From "not give"

Perhaps it is a good time to review briefly how we have arrived at our present position. Because of the R+SSA rule the model needs to be able to infer 'Janet + re-exchange'. Actually, this is not quite correct. To satisfy the R+SSA the model need only make a plausible case that Janet might like the idea of re-exchange. To put this slightly differently, the model needs to infer 'Janet + HYP re-exchange'. (If it is also able to infer 'Janet - HYP re-exchange' the pluses must slightly outweigh the minuses according to our yet nonexistent motivation/action theory.) To this end rule (82) allows the model to infer 'Janet - HYP give present'. Furthermore, rule (69) tells the model that it can get to 'Janet + HYP re-exchange' if it can only get to 'Janet - top'. So this leaves a gap between "not give" and "not want" which we will now attempt to fill.

It seems obvious that the two statements, "giving" and "having" are related, and in particular the latter is a prerequisite of the former. This suggests that we establish a rule relating actions and their prerequisites similar to that for actions and their results.

(86) 'If S is a prerequisite of A, then $P\left\{^+_\}_{-}\right\} A$ is a potential reason for $P\left\{^+_\}_{-}\right\} S$

Note that again the rule has been expressed in terms of potential reasons. The justification for this is the same as before. Just because a person wants to do something does not mean that he will be kindly disposed toward the action's prerequisites. If it is a hot day one may like the idea of baking a cake for
dessert that evening without liking the idea of having the oven on. Or I may, in complaining about a doctor's bill, say that I wish I were on Medicare without meaning that I wish I were 65. If I tell a friend "I would really like to see the show at the Museum of Modern Art, but who can put up with New York in August", my friend must realize that I am giving him a reason why I might want to be in New York while at the same time telling him that it is not a sufficient reason. Furthermore, should I then proceed to go to New York, he should recognize that what previously was not sufficient reason has at least become part of my reason for going.

There are however, several problems with (86). While the positive rule expressed by (86) (+A is a potential reason for +S) is reasonable, the negative rule does not work very well at all. For example:

(87) That I do not feel like making a telephone call does not imply that I have a potential reason for not wanting to be near a phone I can use.

After all, I have been sitting here at my desk for some time, why should I get up just because I do not feel like making a call. The same goes for:

(88) That I have no intention of going skiing this month (June) does not imply I have a potential reason for not wanting to have skis I can use.

(89) (Assuming I am at the library) That I do not want to take out any book does not imply I have a potential reason for wanting to be some place else.

In each of these cases the problem is the same: since it never
occurred to me to go skiing in June (for example), not wanting to do so is irrelevant to my interest in owning skis. This suggests the following revision of (86):

(90) 'If S is a prerequisite of A, then P + A is a potential reason for P + S'

(91) 'If N1 (X is reason for (P + Y)) is true and not X becomes true, then assert not X is a reason for N1 no longer being true'

Rule (90) is simply the positive portion of (86). Rule (91) however differs in two important ways from the negative portion of (86). First, it assumes the existence of the assertion N1. For an example like (92) this requirement is satisfied, as the internal representation would be something like (93).

(92) After Janet bought a top for Jack's birthday she learned that he did not want a top. The next morning she took the top to the store.

(93) 'n1 (Janet + (Janet give Jack top))
    n2 (Janet + (Janet have top))
    n3 (n1 is reason for n2)'

The presence of n3 in the data base is due to rule (90). Note that in the troublesome cases (87)-(89) (where we do not want (91) to apply) there will be no such reason assertion.

Another advantage of (91) over (86) is that by making the presence of N1 the key requirement, (91) can handle cases that (86) could not. For example, (91), but not (86) will handle the examples about the art show, like (24) rewritten here as (94).

(94) Janet took her painting out of an exhibit at a local
art gallery because she decided that the painting was not good enough. However, an art critic convinced her that it was really quite good. That afternoon, after eating lunch, she took the painting to the gallery.

Here the logic would look like:

\[
(95) \quad \neg(\text{Janet believe painting bad}) \quad \text{reason} \quad \neg(\text{Janet + (Painting not in show)}) \quad \text{reason} \quad \neg(\text{Janet + (Janet take painting out of show)})'
\]

Since the painting not being in the show is not a prerequisite to its being considered bad (86) would not apply. Yet, when we learn that Janet no longer believes that the painting is bad we should consider this as reason against her earlier decision.

The other difference between (86) and (91) is more problematic. While (86) allows the model to infer that Janet negatively evaluates the top now that she no longer intends to give it to Jack, (91) only allows the model to infer that she no longer has any reason to positively evaluate having it. Now for some purposes this is enough. For example any reasonable motivation/action theory would conclude from her lack of motivation:

\[
(96) \quad '(\neg(\text{Janet + (Janet have top)}))'
\]

That is to say, the model no longer believes that Janet actively wants the top. Furthermore, given that actions have a cost associated with them, the lack of a reason for doing something is sufficient to allow the model to infer a preference for not
doing it. That is:

(97) 'Janet + (Janet not get top)'

The problem arises when the model needs to infer:

(98) 'Janet + (Janet not have top)'

In contrast to (96), (98) says that Janet actively does not want the top.

Now if the model is told, as in (77) that Janet does not want the top, or if it can infer that she does not want the top because we know (via the take back rule) that she intends to exchange it, then the model can connect her active dislike of the top to (96) by the reasonable rule that anything contributing to lowering a positive desire for X also can be a contributor to a desire for not X. However, in examples like (92) this is not enough. In (92) the model is not told that Janet intends to re-exchange the top. It only knows that she is taking the top to the store and must infer that she will exchange it. To do this the model uses the R+SSA rule, and hence it needs to infer Janet's reason. But (91) does not provide a reason. If Janet would rather not have the top she has a reason; if she does not care one way of the other she does not.

We could solve this problem with a rule like:

(99) 'If P has lost all reasons for P+X then P-X'

But this rule is probably not correct. Consider:

(100) After Janet bought a top to give to Jack for his birthday she learned that he did not want a top. Question: At this point would Janet rather have the top or not?

It would be reasonable to say that she would rather not have
the top, but "I don't know" also seems fine. Even more damaging to (99) is an example like:

(101) Janet went to the library to return a book and to get one on meteorology. After ten minutes she had finished both tasks.

Question: At this point would Janet rather be in the library or not?

Here "Not" seems strange, and "I don't know" much better. The difference seems to be that in (100) getting the top (and hence having it) was a mistake, while in (101) going to the library was not. Such examples suggest that if we want to salvage rule (99) we should reformulate it to read:

(102) 'If P has lost all reason for P+X then (not (P+X)) is a potential reason for (P1-X)'

(103) 'If P judges state X a "mistake" then the potential reason mentioned in (102) is sufficient to be an actual reason'

Rule (102) is a softening of (99) while (103) would be part of the motivation package.

It is hard to know what to make of (102) and (103). In essence they are saying that a person may have a "grudge" against a state of affairs because it was brought about by a mistake. While this may have some psychological validity, it is hard to believe anything so psychological (as opposed to logical) is going on in, say, (92). In taking the top back Janet’s behavior seems to us perfectly reasonable.

Now, we have been trying to fill the gap between '(not (Janet + give top))' and 'Janet + HYP re-exchange' by having
the model infer 'Janet - top' and letting rule (69) take it the rest of the way. What we have found is that while we can infer '(not (Janet + top))', the gap between this and 'Janet - top' is at best problematic and at worst unfillable. There is, however, a way to achieve our goal without inferring 'Janet - top'.

Suppose I would like a particular object Q. It would not normally occur to me to re-exchange W in order to obtain Q simply because to do so would require giving up W. To put this another way, some results of an action can be classified as benefits or costs, and giving up W would normally be a cost. Now since giving up the top is a cost, the derivation of '(not (Janet + top))' means that the cost to Janet of re-exchange has been lessened, hence re-exchange has become more attractive. We might express this idea with:

(104) 'For P, P owning OBJ is generally preferable to P not owning OBJ'

(105) Addendum to rule (69): 'If X is generally unpreferred then (not (P-X)) is also sufficient.
(Rule (69) said that if X is a result of action A, P+X is a potential reason for P+A.) Rules (104) and (105) allow the model to infer 'Janet + re-exchange' without ever inferring 'Janet - top' and hence the problems we faced with rule (94) are not relevant here.

As far as it goes, rule (105) is reasonable enough. However, it should be recognized that (105) leans very heavily on the motivation/action theory, and to this extent is only a partial explanation of how the model should conclude that Janet has reason to want re-exchange.
§3.7 The Representation of Buying a Present

This section will be concerned with one aspect of how to represent the notion that Janet is buying the top as a present for Jack. The discussion will have no particular impact on the rest of the paper so the impatient reader may safely skip to §3.8.

The representation given in (93) is oversimplified. For example, it does not account for the following obvious possibility.

(106) Janet was going to get Jack a top for his birthday. When she heard that he did not want one she decided against it.

Question: Does Janet still intend to get Jack a present?

Answer: Most likely.

The representation we have used so far would simply note that she was no longer going to get Jack a top, but nothing else. We can account for (106) by adopting a representation on the order of:

(107) '[(Janet + (Janet give Jack present))] 
       [ (Janet + (Top be present)) ] 
       imply 
       (Janet + (Janet give Jack top))'

Of course, we could also try to account for (106) by adding a rule to the model which says that if a person decides against getting one present, look for his getting something else. However, this rule seems unlikely in light of examples like:

(108) Today was Jack's birthday. Janet went to the store.
She decided to get Jack a top.

Question: When did Janet decide to get Jack a top?
Answer: While she was at the store.

Question: When Janet decide to get Jack a present?
Answer: I don't know. (or Before she went to the store.)

Example (108) is further evidence that two separate intention statements are needed. Or again:

(109) Today was Jack's birthday. Janet to the store.
She decided to get Jack a top. It was, after all, something she could afford, and it was not too heavy to carry home.

Question: Why did Janet get Jack a present.
Answer: It was his birthday.

Question: Why did Janet get Jack a top?
Answer: It was something she could afford and it was not too heavy to carry home. ("It was his birthday" is also acceptable providing "top" is unstressed.)

Again, the need for the causal statements to refer to different assertions argues for two separate intention assertions. ¹⁸

¹⁸Actually we have only shown that two evaluation statements are needed whereas (107) has three. This suggests that we might do without either the "top be present" evaluation or the "Janet give Jack top" evaluation. At first glance we might be able to do without one or the other, but doing so would probably create more problems than it would be worth. For example, to do without "top be present" would mean that the model would have no assertion corresponding to the notion that Janet chose the top. However it would seem possible to define
That we assume Janet intended to get Jack a present before choosing the present is a fact peculiar to present giving, and does not apply to giving in general. For example:

(I10) Mother had just baked a cake. She gave a piece to Jack.

In contrast to (108) we do not assume in (110) that Mother will give Jack something else if he does not want the cake. Also note that if we substitute "a bite to eat" in place of a piece of "cake" in the questions "When did Mother decide to give Jack a piece of cake?" and "Why did Mother give Jack a piece of cake?" we either get the same answer as with "a piece of cake" or a puzzled look. This reflects our belief that Mother never considered giving Jack anything to eat other than the cake, so if the questions can be interpreted at all it must be that "a bite to eat" is simply a funny way of referring to cake.

---

choose X in terms of having two evaluations, one mentioning X and one not, where the former is a more specified version of the latter. That is, we have '(Janet + (Janet give a present))' and '(Janet + (Janet give a top))' and since the second is a more specified version of the first, the pair constitutes Janet's choosing a top. Then a reason for choosing would be any reason for the specified intention which does not also apply to the less specified intention. However, this scheme seemingly requires a bloated data base as illustrated by a situation in which Janet chooses the top because it is light enough for her to carry. Now lightness is a reason for carrying a top only if there are options about what to carry.
Actually, the difference seems to be that birthdays elicit present giving in general, whereas the situation in (110) suggests that only cake giving was intended. Note that hunger elicits food desire in general and if we preceded (110) with "Mother knew that Jack was hungry" we would get behavior similar to the situation with presents.

§3.8 The Morality of Present Giving

Let us now return to the question of how to formulate the rule which will take the model from "Jack not want" to "Janet not give". As we last formulated it in (82), it said that if a person positively evaluates giving something to someone who does not want it then he has reason to change his mind. While this rule is true for a wide variety of cases, such as:

\[\begin{align*}
\text{Janet} & \quad \text{carry present}
\end{align*}\]
\[\begin{align*}
\text{specifies}
\end{align*}\]
\[\begin{align*}
\text{Janet} & \quad \text{carry top}
\end{align*}\]
\[\begin{align*}
\text{reason}
\end{align*}\]
\[\begin{align*}
top & \quad \text{is light}
\end{align*}\]

This means that the model must have:

But this would seem to require two parallel sets of assertions. The first set states everything Janet intends to do with the top, and then the second set is identical to the first only with 'present' substituted for 'top'. If this is the case then we would obviously be better off with (Janet + (top to be present)) as it would take less space in the end. Trying to do without the repetition of "give" gets us into different problems, which however are similar in that they are inconvenient rather than lethal.
(111) Jack was going to hand Bill an ashtray.
(112) Jack was going to hand Bill a piece of cake.
(113) Jack was going to hand Bill the recipe for the cake.
(114) The official was going to hand Bill his passport.

it is not always true. Consider:

(115) The policeman was going to hand Bill a parking ticket.
(116) George was going to hand his secretary a letter to type.
(117) The official was going to hand Socrates the hemlock.

If we were told (assuming we couldn't figure it out already) that the recipients in these last three cases did not want the thing given to them, we would not want to infer that the givers would no longer intend to give it. The difference between (111)-(114) and (115)-(117) is that with the former, but not the latter we assume that the receiver's wanting the thing is part of the reason for giving. In (111) for example we might assume that Bill had asked for the ashtray. This suggests that we replace (82) with a combination of rules on the order of:

(118) 'Some things are useful to have, and people like owning things'
(119) 'If P1 gives OBJ to P2 and the result is generally favorable to P2 assume P2 + P2 have OBJ a reason for P1 give OBJ'

When this is combined with (91) ('If the reason for doing something becomes false it is no longer a reason for doing it') the end result is that once the model learns that Jack doesn't want a top it will no longer be sure Janet is going to give it to him.
We could improve (119) quite a bit, but it is not worth
doing here simply because (119) is probably not relevant to
our problem. To see this we must first note that (119) is
a low priority rule. Consider:

(120) Jack was spilling ashes all over the carpet. Janet
went to get an ashtray to give to him.

In (120) we do not assume that Janet believes that Jack wants
an ashtray, although ashtrays are inherently useful. This sug-
gests that whatever its final form (119) will have the rider
"unless there is an already established reason". But if this
is the case, (119) will not apply to Janet's giving Jack a top
since there is already a reason for "giving"; it is Jack's
birthday.

Now it might be possible to formulate the rider to (119)
so as not to exclude the case of birthdays, but I haven't tried
because there are other reasons for assuming that the relation
between "Jack wanting the present" and "Janet's giving" is
different than (118) and (119) would predict. First, in any
act of giving where (119) is applicable we can be confident
that the giver believes OBJ is wanted. This is not necessarily
true for presents however. So (121) is odd will (122) is not:

(121) "Pass Alice the salt! I hope she wants it."
(122) "I am going to give Alice a kite as a birthday
present. I hope she wants it."

Passing salt is a typical case where rules (118) and (119) are
applicable, so we assume that no attempt would be made to pass
the salt unless it was already understood that Alice wanted it.
With birthday presents this is not the case. Or again, if Jack
gives Janet his pocket knife we do not know if it is a loan or a gift. If afterwards Jack says "I hope she wanted it" we assume it was a gift since loans follow rule (119).

Most importantly, in the case of presents, but not the other examples of giving something which is wanted, there are moral considerations.

(123) Jack found out Frank did not like the present he had given him, and Jack felt guilty.

(124) Jack found out Frank did not like the piece of cake he had given him, and Jack felt guilty.

While (123) is reasonable (124) is odd. Or again:

(125) Janet was thinking of getting Jack a top for his birthday. "Jack does not want a top," said Penny. "I don't care," said Janet. "Janet isn't a very nice person," thought Penny.

These examples indicate first, that the relation is not explained by rule (79), and second, that the relation has moral overtones.

But can we be more precise? The trouble is that to define the relation between "'give OBJ as present'" and "'receiver want OBJ'" requires a theory of moral relations which does not exist.\footnote{I should say rather that as far as I know such a theory of moral relations does not exist. Since recognizing the moral undertones in example (1) I have been looking at theories of ethics, and while I have found many of the controversies in the field interesting in an abstract sort of way, with one exception, the material has not been relevant to my task. However, ethics is not my field, and there may very well be} Previously we were able to define the relation between
give X to Y and Y have X by saying that the latter was the result of the former. We could use concepts like "result of" and "prerequisite of" because the concepts are so well known that people have no hesitation in accepting them. Of course, even these concepts must prove their usefulness by allowing us to capture the right generalizations, but they are so obvious that they are presumed useful till shown otherwise. Since there are not, to my knowledge, convenient moral relations, we will make some up. But in doing so we will be depending more on intuition than facts, and in particular will not be presenting evidence on how the proposed relations capture the right generalizations.

The relationship needed here is that of moral prerequisite, MPR. So we will say:

(126) 'p1 + (p2 give p1 OBJ) is an MPR of p2 give p1 OBJ as present'

things I could use which I simply have not come across. If the reader happens to know of any, please let me know.

The one exception mentioned above is Searl's derivation of "ought" from "is" (Searl '64). I happened to be familiar with this prior to my work on this paper, and I suspect that his demonstration of how moral considerations may be closely related to non-moral ones made it easier for me to see the relevance of morals to example (1). However, Searl does not go into the specifics of moral relations so his influence is more in style than specifics. Furthermore, it is not necessarily the case that I support Searl against his philosophic critics. They are arguing about the ultimate logical relation between "ought" and "is". From my point of view what Searl has shown is that in a practical sense (and what is pragmatics if not practical) "ought" and "is" are quite intertwined.
The relation is named moral prerequisite because there is an obvious parallel with the usual kind of prerequisite, namely:

\[
\begin{align*}
&\{\text{MPRs generally should be} \} \text{ satisfied before the} \\
&\{\text{PRs must be} \} \text{ onset of the action'}
\end{align*}
\]

(128) 'If action \( A \) is to be performed at time \( T \) and an \( \{\text{PR} \} \) is not satisfied at time \( T \) then
\[
\begin{align*}
&\{\text{generally should} \} \text{ not be done'}
\end{align*}
\]

Because of this close paralleling, the main rule for prerequisites, (90), also applies to MPRs. (This rule says that if \( S \) is a prerequisite of \( A \) then \( P+\)\( A \) is a potential reason for \( P+S \).) So if we generalize this rule to MPRs and apply it to a story where Janet intends to give Jack a top, it would predict that Janet wants Jack to want the top because she is planning on giving it to him. (Hence this rule explains (122).)

(129) Being invited to a party is usually an MPR of going to it

(130) Being sober is an MPR of driving

(131) Having permission is an MPR of using an object which belongs to somebody else.

Although (127) and (128) do not reflect it, the concept 'generally should' is meant to be a relation between a person and a state or action. This relation could be paraphrased "\( P \) generally should see to it that \( X \)." One important property of 'generally should' is:
(132) 'If P generally should X then P has potential reason for P+X'

This rule, along with the motivation/action rules, will allow the model to infer that Jante's giving the top to Jack is biplausible after she learns that he does not want a top.

Actually, one interesting possibility suggested by (132) is:

(133) 'P generally should Y' ≡ 'N1 (P has a potential reason for P+Y) and N1 is inferred from a moral rule'

I do not know if the reduction of 'generally should' to 'potential reason' will work, but besides eliminating a predicate, (133) would also make rules (132) and (127) unnecessary. However, since for the purposes of this paper it will be just as easy to retain "generally should" in the rules, this will be done.

Another property of 'generally should' is its relation to guilt.

(134) If it is necessary to infer for internal reasons²⁰ N1 (P1 guilty about A) then see if P1 did, or intends to do, A and N2 (P1 generally should not do A), and if so then N2 is a reason for N1

²⁰The point of the phrase "for internal reasons" is to allow this rule to be used only when something else in the story suggests that the possibility of guilt should be explored. If an external question were asked about guilt further evidence would be needed, suggesting that ultimately this phrase should be replaced by some aspect of the theory of evidence.
Rule (134) allows the model to understand why Jack feels guilty in a story like (123). However, the logic of guilt is fairly complicated and we have hardly exhausted the subject. One particular case which is important in eventually understanding (1) is the logic behind Janet's guilt in:

(135) Janet gave Jack a top for his birthday. Since Jack did not want a top he asked her if she would take it back. She refused, but she felt guilty about it.

What we need here is some rule which will allow the model to infer that Janet should take it back. Rule (136) will do the job, but certainly needs more work.

(136) 'If N1 (P1 guilt about X) and N2 (P2 (≠ P1) - X) and (P2 - X or - result of X) is a reason for (P2 + P1 do Y) then N3 (P1 HYP should Y) and N1 and N2 cause N3'

One problem with (136) is that it does not handle:

(137) Janet got Jack a present for his birthday which he did not like. That evening he asked her if she would deliver a package which Mother had asked him to deliver, and, still feeling guilty about the present, she did it.

This suggests a weakened form of (136) which does not require that P2's action somehow corrects the action P1 feels guilty about. We get this weakened (136) by deleting the first part of the second line of the rule. My guess is, however, that (136) should be expressed in such a way as to allow both uses, the difference being that we are more sure of the rule when the
connection is there. So, it might be that (136) will allow the model to assume a guilt motivation if the connection between X and Y is present, but if it is not the model must be told, or given further evidence that it is guilt at work.

§3.9 Two Paths to 'Janet + re-exchange' and the Problem of Conditionals

The study of the rules which connect "Jack not want" with "Janet not give" is now done, so we may return to the original question asked in §3.1: in example (1) how is the model to infer a reason for Janet's re-exchanging the top? Actually, rather than immediately tackle (1) let us consider the problem of Janet's motivation in a slightly simpler story.

(138) Today was Jack's birthday. Janet got him a top.

However Jack did not want a top and Janet ended up taking it back.

Even here the answer is complicated by the fact that given the rules we have already established, the model can infer two motivational scenarios for Janet. One goes as follows:

(139) 'From (Jack - top) the model infers Jack - (Janet give Jack top) using rule (69) and the motivation/action theory.

Using (126) and (128) it infers that Janet generally should not give Jack the top.

Using (132) and the action/motivation theory it infers that her giving Jack the top is (at worst)
biplausible (and in doing so assumes that Janet has yet to give Jack the top.)

Using (91) the model infers that her wanting the top is (at worst) biplausible.

Rules (68) and (69) suggest that a potential reason for re-exchange would be (Janet - top), while (104) and (105) combine to say that even (not (Janet + top)) is sufficient to be a potential reason.

From the last two inferences it infers Janet + HYP re-exchange'

To put this in every day terms, having bought the top she found that Jack did not want it and hence she decided not to give it to him. The reason for having the top then evaporated so she re-exchanged it. The second motivational scenario is:

(140) 'Infer (Janet generally should not give Jack the top) as was done in the first two sections of (139).

Rules (68) and (69) suggest that a possible reason for re-exchange is the top's owner not wanting the top. One possible owner is Jack, if the model assumes that the proposed "give" has taken place, hence we may infer that Jack + Janet re-exchange the top.
Jack + (Janet re-exchange the top) is inferred from Jack - top, which together with a) the knowledge that Janet generally should not have given Jack the top, b) she did give it to him, and c) rule (136), enables the model to infer that Janet generally should re-exchange the top.

This plus either rule (132) or definition (133) enables it to infer that Janet has a potential reason for re-exchanging the top.

Again, putting this in everyday terms; Janet probably felt a little guilty about getting Jack something he did not want, so to help out she should re-exchange the top for him.

The construction of one or perhaps both of these scenarios enables the model (and I would assume a human reader also) to understand the connection between events in a story like (138). However, (1) is a more complicated story than (138) and presents two problems which the latter does not have. First, we are told that Jack will make Janet take the top back. This obviously places additional constraints on what the scenario can look like, and in fact makes one of them, (140), more likely than the other. The problem is deciding exactly what the constraints are, and how they interact with our other rules to make (140) seem more reasonable. This problem will be discussed in §3.10 and §3.11.
The second problem is that Penny's last line in (1) is really a conditional statement. That is, she is really saying "If you buy the top for Jack and give it to him then he will make you take it back". While I do not intend to go into the problem of the conditional nature of "He will make you take it back", conditionals being a major research problem in their own right, I do want to indicate how the general problem of conditionals manifests itself in this particular instance.

Consider our state of knowledge just before Penny utters her last line. She has already told Janet that Janet should not get Jack a top, and has indicated why, so we realize that there are reasons why Janet should not get Jack a top, and we are no longer sure that she will. The first problem is for the model to realize that Penny's statement is a conditional. If we know that "it" in "take it back" refers to the present, this is reasonably easy since Janet would have to have bought the top before it could possibly be taken back, hence unless we assume that the sentence is conditional the sentence is a non sequitur. However, it is by no means clear how the model is to decide that "it" refers to the present. Not only is there another possibility, the top Jack already owns, but if the model uses a simple rule like "'it' refers to the last mentioned inanimate object" the model will decide on the wrong referent. So the problem of deciding that Penny's statement is a conditional becomes the problem of choosing the right referent, a notoriously hard problem.

The second problem under the heading of conditionals is as follows. Prior to the last line we know that there are
potential reasons why Janet should not get the top, namely that Jack does not want one. Penny, in hypothesizing Janet's getting the top, must also be making some assumption about why it is that Janet is not heading these potential reasons. There are several possible assumptions that Penny could be making.

(141) It is not true that Jack does not want a top.
(142) Janet does not care if Jack wants a top or not.
(143) Janet at the moment does not believe Penny.
(144) Penny never warned Janet.

However, (141) cannot be Penny's assumption, or else there would be no sense to her statement that Jack would make Janet take the top back. As for (142), it is not compatible with scenario (139) since (139) depends on Janet's deciding not to give Jack the top because he does not want it. Hence any person who interprets (138) say, along the lines of (139) cannot be choosing (142), and hence the model of such people should not choose it either. The problem for the model then is selecting (143) or (144) instead.

§3.10 Defining "make"

In (1) the effect of "he will make" on our understanding of the last sentence is to have us assume a scenario along the following lines.

(145) Janet gives the top to Jack and he in turn tells her that he does not want a top. Janet has no particular intention of re-exchanging the top for him but he asks her to and she feels compelled to do it.
It is clear that of the two original scenarios, (139) and (140), (145) most closely follows scenario (140). The major point is that in both (140) and (145) we assume that Janet actually gave Jack the top. Also, although we did not mention it, in (140) the model must assume that Jack gave his permission to Janet to have the top taken back. (When we infer "Janet generally should re-exchange" the model must check to see if any of the MPRs of re-exchange are violated, which would count as reasons why she should not do so. One MPR of re-exchange is that the actor has the permission of the owner, in this case Jack.) But (140) is mute on one crucial point, did Janet offer to re-exchange the top before Jack mentioned it, or did Jack ask first? By adding "he will make" we decide the issue. What we would like to know then is what is the information activated by the presence of "make" which allows us first to select (140) over (139) and then to further specify (140) so it becomes (145).

Naturally, the first place to look is the meaning of "make". "Make", of course, can appear in many different settings, but for the purposes of this discussion I will pretend that it only appears in the setting:

(146) ANIMATE make ACTION

As a first attempt at a definition, clearly inspired by examples like (1), (147) and (148), I came up with (149).

(147) The robber made the bank teller hand over the money.

(148) Mother made Jack wash his hands.
(149) P1 make P2 do A2
     'P1 inform P2 that P1 + (P2 do A2)
     P1 inform P2 (or P2 already knows) that if P2
does not do A2 something bad will happen to P2
     The net result is that P2 does A2'
The trouble with such a definition is that it does not take into
account cases like:

(150) My father made me cough by pounding me on the back.

(151) My father made me get a job by cutting off my
     allowance.

(152) The pedestrian made me jam on the brakes.

In none of these cases does P1 inform P2 of anything, and in
the latter two we cannot be sure that P1 intended P2 to do A2.
Examples like these suggest a really stripped down definition of
make, such as:

(153) P1 make P2 do A2
     'P1 does some act A1 which causes P2 do A2'

Before considering the problems with (153), let us dispose of
the possibility that "make" is really ambiguous (that is "make"
as it appears in (146) and definition (149) corresponds to one
meaning and (153) to another.) First, note that (153) can be
brought closer to (149) in net effect if the model has an in-
fERENCE rule like:

(154) 'A common way to cause someone to do something is
to inform him that you would like it if he did it'

Furthermore, (154) is needed independently of definition (153)
since we also want to be able to handle situations like:

(155) Jack got Janet to re-exchange the top.

(156) I don't know how Jack did it, but Janet actually re-exchanged the top for him.

In both cases we assume that Jack asked her to do it, although in (156) there is an implication that extra leverage was needed. The ambiguity theory of "make" would cause us, in effect, to have two copies of (154) around, one for use in cases like (155) and (156), and the other hidden in the second definition of "make", (149). 21

Another argument against the ambiguity theory is based on the observation that the two theories make different claims about a sentence like:

(157) My father made me get a job.

If the ambiguity theory is correct, then (157) is ambiguous, otherwise it is simply unspecified as to Father's action.

A linguistic test for the distinction between ambiguity and unspecificity is afforded by the following examples.

(158) Jack is under the table and Joe is under the chair.

(159) Jack is under the table and Joe, the chair.

(160) Jack is under the speed limit, and Joe, the maximum weight for light weights.

21 Furthermore, I suspect that (154) is really the same rule as (60) only expressed slightly differently. If this is true, then the examples supporting (60) also support (154).
(161) Jack is under the table and Joe the speed limit.
The point here is that in examples like (158) where the second clause is interpreted as having material from the first (or has had material similar to the first deleted from it, depending on how you look at it) one gets strange sentences if the readings do not stay constant between the two clauses. (I am assuming of course that "under" in "under the table" and in "under the speed limit" has truly different readings. So in an example like (161) the first clause has one reading, the second requires another. A second illustration of the same point:

(162) Fred will beat the wall, and he will do so to the table also.

(163) Fred will beat the wall and he will do so to the other contestants in the race also.

Naturally, in cases where we have unspecificity, rather than ambiguity, there is no such phenomenon:

(164) Jack kicked Bill and he did so to Fred also.

There is no reason why this could not be describing a situation where Jack kicked Bill with his right foot, and Fred with his left.

Getting back to "make", consider a situation where my father has made me get a job by cutting off my allowance (but he never told me to get a job, and perhaps was even unaware that his action would cause me to do it) and he also told my sister to get a job (with an implied threat). It is perfectly reasonable (although slightly awkward) to give a partial description of the situation as follows:
(165) My father made get a job and he did so to my sister too.

Again the conclusion is that make is not ambiguous (at least not in the way claimed).

Having resolved to make to with one definition for "make" in contexts like (146), it is clear that (153) is not the definition we need. (Pl does something causing A2). For example, consider:

(166) Janet walked into the garden. Jack was there fixing the fence. Seeing her, he said, "Janet could you hand me that hammer?" Janet did so.

In (166) Janet cannot say that Jack made her hand him the hammer, although it is clear that his request caused her to do it. Hence (153) is insufficient since it would allow the use of "make" to describe (166). In fact, when one starts looking closely, the use of "make" is quite bewildering. To give only one example, (167) I met Boswell, the notorious rich eccentric, on the street the other day. He offered me a hundred dollars if I would raise my right hand. Naturally I did so.

I would not say in (167) that Boswell made me raise my right hand. But note that Boswell could, quite legitimately, claim that he had made me do so.

The best definition I have come up with so far is the following:

(168) Pl make P2 do A2

1) 'Pl does some act A1 which causes P2 do A2'

2) 'P2 did not have free will in doing A2'

In (167) we would say that I retain the belief that I had free
will in doing what I did. Boswell, knowing in advance that he could get me to do it believes that I did not really have free will, and so he made me do it. The trouble with (168) is that the concept of free will is so vague that (168) is hard to implement. However, it does seem to fit reasonably, and given the complications of "make" the vagueness of "free will" perhaps is in order. At any rate, most of the further analysis will only depend on clause (i) of (168) and any work based on (ii) will be clearly labeled.

§3.11 How "make" Influences the Understanding of (1)

In §3.9 we noted that if all we were told were that Jack did not want a top and Janet ended up taking it back, there were two plausible scenarios which would explain her action. In (139) it was assumed that Janet, presumable learning of Jack's antipathy towards tops, decides not give him the top, and, no longer having any use for it returns the top to the store. In (140) on the other hand it was assumed that Janet does give Jack the top and presumably feeling somewhat guilty about giving Jack something he does not want, Janet re-exchanges the top. But we then noted that in (1) where we have "He will make you take it back" we do not get scenario (139). Furthermore, the scenario we do get, (145) is more specific than (140) in that it specifies that it was Jack who initiated the idea of re-exchanging the top, not Janet.

Roughly speaking then, our task is twofold. First we want to show how the model is to choose (140) over (139), and then how it is to elaborate (140) to arrive at (145).
The main difference between (139) and (140) is whether Janet is assumed to have actually given Jack the top. The other differences between the two, in particular Janet's motivation, follow from this choice. For example, if Janet has not yet given Jack the top then she is its owner and the model needs a reason why she might not want the top, which it finds in her loss of interest in having it.

If we can show then how "make" leads the model to the assumption that Janet did give Jack the top we will have shown how the model chooses (140) over (139).

To break into this topic as easily as possible, we will start out by considering how to choose (140) given Penny saying "Jack will make you re-exchange the top at the store." This is a case where there is little problem in deciding exactly what it is that Janet will be doing. Then to my mind, the most plausible path to the choice of (140) goes as follows:

(169) From (168) the model learns that Jack will do something to cause Janet to re-exchange the top. Using (154) it decides that he probably asked her to do it. From the nature of "ask" (which we have not gone into) it can conclude that he probably wanted her to re-exchange the top. In trying to decide why he would want her to do it the model looks at re-exchange and from (68) and (69) concludes that one possibility is that Jack does not want the top's owner to remain the top's owner. But in trying to find out who the top's owner is it finds that it might be Janet (if "Jack's telling" occurs before...
"give") or it might be Jack (if "telling" occurs after "give"). It then notes that it already knows that Jack does not want Jack to be the owner of the top, whereas it does not have a ready-made reason for why Jack would not want Janet to be the owner, so it chooses Jack as the top's owner, and hence chooses "tell" as following "give".

The key step in (169) occurs in the choice of Jack over Janet as owner of the top. Now this is not an inevitable choice. Janet could possibly be the owner, and it would not even be all that difficult to think up a reason why Jack does not want her to be the owner (he is afraid that her being the owner will lead to his being the owner). But it is also true that while "make" causes us to choose (140) it is not an inevitable choice, and we could imagine Jack telling Janet before she gets around to giving him the top. It is just that everything follows more directly if we assume that Jack owns the top. Hence (169)'s lack of necessity is not a strike against it.

There is however another possible path, but it requires clause (ii) of (168) (the definition of "make"). Clause (ii) is the "free will" clause. If we accept this clause as correct, the model could have a chain of inference as follows:

(170) By (168) the model knows that Janet does not have free will in her decision to re-exchange the top. It tries to figure out how this might be the case and notes that one way is that Janet believes that something bad will happen to her if she does not do what Jack requests. By (136) the model can
infer that Janet would feel guilty if she refused to do something Jack wanted her to do, but this is only true if she has actually given him the top (as a rule (136) requires doing something you should not have done).  

However, not only does (170) require clause (ii) of (168), but it is also on shaky grounds in its use of (136). Rule (136) is the rule which says that if you feel guilty about what you did to somebody then you should make it up to him. The strong version of this rule further specifies that what you do to make up for the guilt is to be related to what you did to deserve the original guilt. When (136) was introduced it was suggested that perhaps the strong form is required when the story does not mention guilt, but rather the guilt has to be inferred from the guilty party's acquiescence in helping the non-guilty party. If on the other hand the story mentions guilt then we need no further evidence (in the form of a connection between what was

\[\text{We need clause (ii) because Jack's just asking Janet to re-exchange the top is not sufficient to cause a chain like (170). In particular, the model might try to figure out why Janet would do what Jack asked her to do, but one possibility would be that she already had a reason of her own for doing it, as in scenario (139). Of course, if Jack asks Janet to re-exchange the top we do feel that (140) is the correct scenario, but this can be explained by the chain of reasoning in (169).}\]
done to cause the guilt and what to alleviate it) to draw the inference that guilt was a motivating factor. Naturally, if this is correct, (170) must use the strong form since (1) does not mention guilt. This in turn implies that at some point in (170) the model must make a connection between Jack's problem with the top and re-exchange. However, this is exactly what scenario (169) does. This means that before scenario (170) can be completed (169) must be completed, leaving scenario (169) as the only way to infer that Janet did give Jack the top.

This argument rests, of course, on the correctness of the hypothesis about the strong and weak forms of (136). While there is a little evidence to support it, the hypothesis can hardly be considered secure. Nevertheless, it is fair to say that scenario (170) seems much less robust than (169).

There has been a hidden assumption in how (169) and (170) were presented, namely that the presence of "make" guides the reasoning process so as to lead to the creation of scenario (140) without (139) ever being considered. There is another logical possibility however. It is possible that both (139) and (140) are initially created and "make" then "chooses" (140) over (139). Now if it is indeed possible to parlay the information from "make" into the ability to go immediately to the correct scenario, it would naturally be preferable to do so, if only from an efficiency point of view (why create two scenarios instead of one). Furthermore there seem to me to be other reasons for preferring this manner of doing things. Consider the case where Penny's last line is
"you will end up taking it back." If one were to ask people what Penny meant by that, I suspect that some would produce a chain of reasoning like (139), others (140), but nobody would mention both. Even if true it is dangerous to take such observations too literally, but the simplest explanation of this phenomenon (and hence the explanation to be preferred) is that in their model of the story people only construct one of the two possible chains of reasoning. If this is true, then it would be odd if the simple change to "Jack will make you take it back" suddenly causes people to produce both chains, only to remove one of them. (Note that it is not the case that people can produce one chain, and then, if it is the "wrong" one go back and produce the other, since, as was pointed out earlier, it is quite possible to produce a coherent story based on scenario (139) where Jack makes Janet re-exchange the top. That is to say, scenario (139) is not wrong, but only less preferred.) So, based on what we have seen so far the "pre-selection" method (using the information from "make" to guide us directly to the best scenario) is to be preferred over the "post-selection" method (creating both scenarios and then using the "make" information to choose between them).

I mention the option of post-selection because while the pre-selection reasoning presented in (169) works quite well if the last line of the story is "He will make you re-exchange the top" or "take it back", there are problems if the last line is "He will make you take it to the store" (as in example (18)). As was noted earlier, although this line sounds odd in the context of story (1), it is understandable, hence any theory of "make"
which accounts for our choice of (140) should do so also when the last line is "make you take it to the store". The problem is that (169) assumes that it is known what Janet has been made to do. But if we have "take it to the store" the model can only be sure of what Janet is doing after it has figured out why she is doing it (that is, it uses the R+SSA rule). The problem then becomes a complicated one of how much knowledge one can assume at any point in the inference chain. While I believe that (169) can still be carried through in the case of "take it to the store", to argue the point would be difficult and probably unconvincing given the general lack of specificity about when and how deductions are made in the model. The point of bringing up the post-selection method is that the information used by the pre-selection method can also be used by the post selection method should the former prove untenable.

Finally we want to show how "make" allows the model to elaborate on (140) to produce (145). The problem is to recognize that when Jack tells Janet that he would like it if she returned the top, he is instigating the action, and not simply giving his blessing to an idea Janet already thought up. To put this another way, Janet did not intend to re-exchange the top prior to Jack's action. Now "make" tells the model that P1 did something (A1) which caused P2 to do A2. However, in general, there may be several causal links separating A1 from A2 which "make" obviously does not know about. In particular, in the case of actions which are normally instigated by conscious decision, we normally assume that to cause P2 to do A2 is really to cause P2's intention of doing A2. This suggests we need
a rule, or set of rules which does the following:

(171) 'If A1 cause (P2 do A2), where A2 is an intentional act, then assume A1 cause (P2 INTEND A2) unless there is a reason to believe P2 INTEND A prior to A1'\(^{23}\)

(I am assuming that an integral part of the semantics of the causal relation being used here is a rule that if an action causes a state, then the state was not true prior to the action. This distinguishes 'cause' from the concept underlying "reason" as this word has been used in the paper. For example, it is quite possible to gain a new reason for wanting to do something you already want to do.) Since prior to the last line of (1) the model has no reason to believe that Janet intends to re-exchange the top, it assumes that she did not so intend. This inference is the important step in going from (140) to (145).

§3.12 Summary of §3

To some degree the ordering of topics in this section has reflected the path I took in thinking about them. While this does give some idea of the problems involved in doing this sort of research, it does not often lead to the most logically ordered presentation, so I will conclude by summarizing the results in a more rational order.

\(^{23}\)We have to check for P2 INTEND A2 because of examples like: Jack was about to stand up when the robber pointed a gun at his head and made him stand up. Here we have one action causing another, not by causing the intention, but by being so much more important than any other reasons for having the intention.
1) The fact that Janet's returning the top makes more sense than, say, Penny's doing it means that the model must have a rule or set of rules which connect buying the top with exchanging it. Initially this can be viewed as two rules; one connecting "Jack not want" with "Janet not want", and the second connecting "Janet not want" with "Janet re-exchange". (§3.1)

2) Looking at the first of these two rules, from Jack's not wanting the top the model should directly infer the inadvisability of giving it to him, as opposed to the inadvisability of Janet's buying it, or having it. (§3.4)

3) Furthermore this rule can only say that Jack's not wanting the top suggests that Janet not give it to him, but she still might and hence "Janet will give" is biplausible. That is, we can reasonably believe that it is true or false. (§3.5)

4) The relation between "not want" and "not give" is, in this instance peculiar to present-giving, and is not true of giving in general. (§3.8)

5) It also seems to be a moral relation because of such things as our negative opinion of Janet if she does not care if Jack wants a top or not. (§3.8)

6) Because of its similarity to the prerequisite relation, I call the rule connecting "want" and "give" a moral prerequisite. However, given the lack of knowledge about moral relations, this should be taken with a grain of salt. (§3.8)

7) To use the rule mentioned in 6 the model must still connect
Janet's not giving the top to her not wanting it. So we infer that if Janet no longer intends to give Jack the top then she has no reason for wanting the top. The rule responsible for this inference does not directly use the fact that "have the top" is a prerequisite of "give the top", but instead is expressed in terms of one desire (giving the top) being a reason for the other desire (having the top). This in turn is inferred from the prerequisite relationship which holds between them. (§3.6)

8) As for the second rule mentioned in 1; since "owner no longer owning object" is a result and cost of re-exchange, the owner not particularly wanting the object is a potential reason for re-exchanging it. (§3.2, §3.6)

9) Steps 7 and 8 taken together form a reason for Janet's re-exchanging the top if she has not yet given it to Jack (since she is not planning on giving it to him she has no use for the top). (§3.9)

10) Another chain of reasoning leading to Janet's exchanging the top is that since she should not have given Jack a present he did not want, she should do something to make amends. (§3.8) As in step 8 we can assume Jack has reason to want the top exchanged and if he does then Janet should do it. (§3.9)

11) While Janet might exchange the top for either of the reasons presented in 9 and 10, by saying "Jack will make you take it back" Penny makes 10 more reasonable than 9. This is due to the definition of "make" plus other rules of inference. (§3.10, §3.11)

12) Serving as foundation for the discussion were the concepts
of positive evaluation (§3.2), common sense theory of evidence (§3.4) and motivation/action rules (§3.4)

13) Not used directly in the above plan, but included because it seemed generally relevant were a discussion of the representation of "Janet will give Jack a top as a present" (§3.7) and a discussion of the problems introduced by the fact that "he will make you take it back" is really a conditional statement (i.e., there is an implicit "if you get Jack a top"). (§3.9)
§4 HANDLING "TAKE BACK"

§4.1 The Old Take Back Rule, and the New

In §2 we devised (19), rewritten here as (172), to handle "take back":

(172) The Take Back Rule (Version I)

'PERSON take back OBJECT to LOCATION where he got it implies that he will undo what he was up to when he got OBJECT originally'

That some rule for take back is necessary is shown by pairs of examples like (14) and (15), or (16) and (17) where replacing "take back" with "take to the store" changes the meaning of the sentence. That the rule looks like it does is due to considerations like:

(173) The rule must check that OBJECT was obtained at LOCATION since in examples where this is not true, (12), there is no inference that the obtaining is undone by the take back.

(174) The rule should talk in terms of "undoing the what he was up to" so that it can handle examples like (8) ("take back the frog") and (9) ("take back the wagon").

(175) An equivalent to the phrase "what he was up to" is needed because "take back" does not simply imply undoing 'get' but rather implies that a whole complex of activities are to be undertaken. For example, we noted that taking back the top was not simply giving the top to the storekeeper, but also getting something in return. Nor does "take back" simply
mean, in this case undo the trade since what is obtained from the store in return need not be what was originally handed over.

It would seem that any successful take back rule must follow (173)-(175). However, as we shall see, there is still quite a bit of room for differences. In this section we will first develop a rule based on (172), namely (176).

(176) The Take Back Rule (Version II)

'This rule is meant to handle PERSON take back OBJECT to LOCATION where the act of taking is N1.

1. This is a medium confidence rule for establishing a superact\textsuperscript{24} of N1.

2. See if PERSON2 obtained OBJECT at (or from) LOCATION (=N2) (that is, LOCATION can be a person, place or institution).

3. Find the highest superact of N2 (=N3) such that PERSON have OBJECT is a result of N3.

4. N1 is a subact of undoing N3.'

In §4.1 we will justify many of the differences between (176) and (172). Finally, in §4.3 we will look at some small problems with (176) and from them a lot of conjecture build what to my mind is a much more intuitively reasonable rule, but one for which there is little evidence at the moment.

\textsuperscript{24}Superact is the inverse of subact. Y is a subaction of X if in the course of doing X one does Y or if doing Y satisfies a prerequisite of X.
§4.1 Justifying Rule (176)

Essentially what part 1 of (176) says is that if you already know why the object is being taken back, then don't use the take back rule to figure our another reason. We need this to account for examples like:

(177) Jack caught a frog at the swamp and took it home. The next day he wanted to make sure the frog got its daily ration of flies, so he took it back to the swamp.

(178) Janet bought a purse at the store. The next day she decided that she wanted gloves to match the purse, so she took the purse back to the store.

In both cases we are virtually told what the superact is, so there is no need to use the take back rule. Note that there can be cases where it is possible to infer a very tentative superact which the take back rule can override (as in (16) vs. (17)).

A slightly different example where we want to inhibit the take back rule is:

(179) Jack got a frog at the swamp. The next day he went to the swamp with the frog. He brought the frog home at lunchtime and then took it back in the afternoon.

While we do not interpret (179) as implying that Jack will get rid of the frog, we do not know why he is taking the frog to the swamp. However, if we had a rule like (180) then (179) could be handled just like (177) and (178) since we would already have a superact.

(180) If a person repeats an action which he has done recently then he did it for the same reason as last time.
The advantage of handling (179) with a rule like (180) is that (180) might be needed anyway. The disadvantage is that (180) may be quite difficult to formulate correctly. For example, it must be restricted to "distinctive" activities, as opposed to common ones like going out of the house. (One possibility in this regard is only considering activities which qualify as significant subactions.)

There is another, reasonably simple way to handle (179). We could modify our take back rule as indicated in (181).

(181) In establishing 'obtain at LOCATION' the take back rule only considers the last instance P2 was at LOCATION'

However, compared to (180), (181) is quite ad hoc so I will assume some rule like (180) will take care of (179). In either case it is clear that (179) does not present great difficulties to the take back rule.

The second change to note in (176) is that it no longer requires that the person who gets the object is also the person who takes it back. We need to do this in order to account for stories like:

(182) Today was Jack's birthday. Janet and Penny went to the store. They were going to get presents. "I will a top," said Janet. "Jack will take it back," said Penny.

(183) Jack bought some milk at the store. The next day his mother took it back.

(184) Jack bought a top at the store. The next day Bill took it back.
In (182), while Jack eventually will get the top, it is Janet who gets it from the store. Jack simply gets it from Janet. Yet we understand take it back as to the store rather than to Janet. In (183) while Jack got the milk, Mother is taking it back, although we recognize that Mother has the "right" to do this. In (184) my informants tell me that they still assume Bill will re-exchange the top, only here they are forced to invent a justification for his doing so, like Jack asking him to do it. What this seems to indicate is that the take back rule cannot specify that the getter and taker must be the same person. That they usually are is presumably due to other considerations, such as: one should not re-exchange, or otherwise affect the ownership status of something one does not own unless one has permission.

The third change to (172) proposed in (176) is in line 4.

(185) Find the highest superact of N2 (=N3) such that PERSON have OBJECT is a result of N3.

Informally, the idea behind subacts and superacts is that people have a hierarchy of goals. If X is higher in the hierarchy than Y it is a superact of Y.

In the first version of the take back rule we used the phrase "what PERSON is up to". (185) is meant to capture this idea more formally. Since "what PERSON is up to" is itself a vague concept, it is hard to show that "superact" amounts to roughly the same thing. However, there are two aspects of (185) which can be shown to be necessary if the rule is to work properly. First, is the need to specify that we want the 'highest' super-
act. Consider:

(186) One day Jack went to the swamp to catch a frog and Bill went along. They both saw a frog and ran after it. Bill grabbed it and handed it to Jack. The next day Jack took it back.

In this story we understand that Jack is taking the frog to the swamp, not to Bill, (185) would account for this since intuitively getting the frog from Bill is part of getting the frog at the swamp. (There is some independent evidence for this intuition. If before Jack returned the frog someone asked him where he got it he would say "the swamp" rather than "from Bill". In fact, such responses even offer the possibility that (185) can be completely removed from the take back rule and incorporated in the rules needed to answer questions about where PERSON got OBJECT.) Note that we cannot account for (186) by saying that "take back" only applies to locations, and not to people. For example we have:

(187) In class one day Jack traded his pocket knife to Bill for Bill's top. Later Jack decided to take the top back.

Here people understand Bill, not the class room as Jack's destination.

As for the requirement that 'PERSON have OBJECT' be a result of N3, this is needed because of a technicality in the definition of subaction. For example, if I catch a frog at the swamp in order to paint him orange, getting the frog satisfies a pre-requisite of 'paint frog' and hence is a subaction of 'paint frog'. However, 'paint frog' is not what is undone by take back,
but rather 'catch frog'.

4.3 Speculation on the Nature of "take back"

While the rule (176) works reasonably well, as a whole it seems rather arbitrary. One is tempted to ask why English needs a word that works in such a strange way. In this section I will theorize as to what is really going on with "take back". However, since the theorizing will not be accompanied by much evidence, and since the ideas will not be very precisely formulated, this section could perhaps be more truthfully called "directions for Further Research".

First, let us again consider the last example, (187). Although I have been keeping away from the question of reference, in this case I will make an exception and ask how it is that the model should choose 'Bill' over 'classroom' as Jack's destination in his attempt to take the top back. As far as I can tell, the only information which makes Bill more plausible that classroom is the knowledge that taking the top to Bill could be part of exchanging it while taking the top to the classroom has no particular significance at all. That is to say, in this case even though the take back rule applies we must still make use of the fact that taking the top Bill is an SSA of exchanging it with him. Furthermore, note that the take back rule only seems to work in situations where "taking" is an SSA of "undoing get". This naturally leads to the speculation that perhaps the take back rule is really dependent on this SSA information in all cases.

To strengthen this speculation, consider the following example:
Janet bought a doll. When she got it home she found that its arm was loose. The next morning she took it back.

Most people interpret (188) as 're-exchange', but some assume that the doll is being taken back in order to be repaired. Or consider:

Mr. Jones bought a new car. After a few days the heater stopped working and Mr. Jones took the car back.

I would assume that most people would interpret (189) as 'fix'. On the surface examples like (188) and (189) are problematic for rule (176). The only way to accommodate (176) to these examples is to assume that prior to applying the take back rule the model has already made the link between '"broken"' and '"return to fix"'. If this is the case then the take back rule never applies (by clause 1 of (176)). However, in light of the earlier speculation, there is another explanation of (188) and (189). Simply put, "take back" does not in its own right imply re-exchange. Or, more precisely, "take back to LOCATION where OBJECT was obtained" does not imply undoing the act of obtaining OBJECT. Rather, in those cases where we have for independent reasons the belief that taking to LOCATION is an SSA of re-exchange the use of "back" reinforces this belief.

If, on the other hand, we believe that taking OBJECT to LOCATION is an SSA of '"fixing"', "back" reinforces that belief. More, generally, I would speculate that if it is possible to infer X from a sentence without "back", and X itself implies (or makes it reasonable) that OBJECT has been at LOCATION previously,
then "back" will reinforce the inference X. So, when the model tries to interpret "take" as "re-exchange", it notes that "re-exchange" itself implies that OBJECT was bought there, and hence this inference is strengthened by the presence of the word "back".

Finally, it may be the case that other words work this same way. Consider the use of the word "again" in:

(190) Janet took her painting out of an exhibit at a local art gallery because she decided that the painting was not good enough. However, an art critic's praise convinced her to take it to the gallery again.

Note that this story sounds fine, while earlier we saw that the same story, missing "again", sounded odd, (23). In the model we might explain why both "back" and "again" help (23) by postulating the use of rule (180). Rule (180) says that if a person repeats an action in a short period of time, he is doing it for the same reason the second time as for the first. Hence the successful application of this rule implies that the action is occurring again. Hence "again" reinforces rule (180) just as "back" reinforces the inference that take to the store is part of re-exchange.

The speculation then is that "back" and "again" have limited meanings which reinforce other inferences. In some cases like those we saw in §2 ((14) vs. (15), and (16) vs. (17)) this reinforcing is sufficient to allow an inference to go through which otherwise would not. To put this in other words, "back" is like
"but" in that it reinforces other inferences rather than being an instigator itself. It differs from "but" in that it contains more semantic information and hence more power over the inferences we make.
§5 CONCLUSION

The factual conclusions of this paper have been summarized elsewhere (§1.1, §2.3, §3.12, §4.1), so there is no need to do it again. Rather I will conclude with some general comments on what I have tried to do here.

I began this work after an off hand comment Marvin Minsky made to me while I was finishing up my Ph.D. thesis. Since example (1) was a prominent example in the thesis, why not give a more precise description of the knowledge needed to understand it? I thought it was a reasonable idea and spent about a day and a half before deciding that it was more complex than I realized, so I shelved the problem. After the thesis was finished I took it up again, only to become bogged down again after a month and a half. At that point I knew that the word "back" was important since (1) sounded horrible without it, but I did not know how "back" worked and the discovery of "backless" stories which sounded OK left me even more confused. Why should "back" sometimes be required and sometimes not?

The people to whom I showed my results agreed that the problem was a tough one. One went so far as to suggest that I was bound to fail. It was his contention that the state of the art would not support an even partially complete account of what went on in our understanding of a single line of a story. Given the number of holes in this analysis, he might be correct. Yet it is my feeling that the results here are in keeping with the original spirit of the inquiry.

Be this as it may, the important question is whether this
sort of thing is worth doing. As I said in the beginning, I am combining linguistic methodology with assumptions from artificial intelligence AI, so it is these two groups I most wish to convince. I do not know what a linguist will make of this paper, but the criticism I have received from people in AI has been reasonably uniform. What is needed, it is said, is a theory of understanding, and a small number of rules investigated in detail will not help. As one more philosophic fellow put it, history shows that trying to induce theory from data is not the way to do things. What he meant, of course, is that ultimately I want to induce a theory of understanding from the particular rules I come up with.  

While I think that it is fair to say that history does indicate that one should not collect data with the hope of somebody else using it, (though Priestly and Brahe are exceptions even here) this is not exactly the situation at hand. While, once established the rules formulated here could be considered data for the next stage theory, right now they constitute a low level theory in themselves. This theory hopefully explains the real data, my intuitions about stories. And there is another good rule to follow in doing scientific research: work on problems which have a reasonable chance of being solved. What then are the doable alternatives to the kind of analysis presented here?

\[25\] Obviously this would not be a criticism I would expect from a linguist, since in the domain of syntax this is roughly what linguists are trying to do. (Or to be more accurate, any criticism of me on this score is bound to hit them too.)
The AI literature has two different answers to this question. One possibility is to attempt a complete theory of reasoning and language, recognizing that no one part is likely to be remotely adequate to explain language use. The other is an exploration of one of the "general" problems, like evidence, motivation, conditionals, negation, time and space, etc.—that is to say, those problems which I have explicitly avoided in this paper. Originally I had intended to conclude with an analysis of these options, but I have decided against it. In the case of the complete theory approach, I still believe that it is not the proper moment for an all-over attack. At its best it can provide an illuminating "suggestiveness", but it seems to me that we now know too much to make "suggestiveness" useful, yet much too little to give an all-over theory any hope of being accurate. As for the "general problem" approach, I now have a new respect for it. The number of times in my analysis where I had to leave off because a general problem got in my way has made me upwardly evaluate the importance and tractability of such problems, in the sense that they now appear to me to be much closer to the "surface". Nevertheless, I think that such studies must be couples with detailed studies of specific cases if they are not to be vacuous.

In the final analysis, of course, there is no sure way to predict which of these approaches will prove most fruitful. You pays your money and you takes your choice.
REFERENCES

Charniak, E. Toward a model of children's story comprehension. AI-TR266, MIT Artificial Intelligence Laboratory, 1972.


Searle, J. How to derive "ought" from "is". *Philosophical Review*, 1964, 73, 43-58.
