# JACK AND JANET IN SEARCH OF A THEORY OF KNOWLEDGE

Eurene Charniak
Artificial Intelligence Laboratory
Massachusetts Institute of Technology

## Abstract

In order to answer questions about children's stories one needs a great deal of "common sense" knowledge. A model Is presented which gives a rough organization to this knowledge along with specifications as to how the Information will be accessed. This rough model Is then used as a basis for tight arguments about narrow Issues (primarily using examples concerning piggy banks.) The paper Is intended as an illustration of how one might go about constructing a theory of knowledge.

## Acknowledgements

This paper is based on portions of an KIT Ph.D. thesis submitted to the department of Electrical Engineering. The thesis Is reproduced as AI Technical Report,266, As In my thesis I would like to thank all the people at the MIT Artificial Intelligence Laboratory who listened to and argued with me on many occasions.

The work reported herein V/PS conducted 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.

## 1 Introduction

Let us consider the problem of constructing an abstract model of story comprehension. To determine what the model, or program, has "understood" about what it has read, we will ask It questions. So a typical story might start:

(1) Janet needed some money. She got her piggybank (PB) and started to shake it. Finally some money came out.

Some typical questions would be:

(2) Why did Janet get the PC?
(5) Did Janet get the money?
(4) Why was the PB shaken?

Questions (2) - (4) are not answered explicitly In the text. That Is, the story did not say "Janet got her PS because she ..." The story does not even contain a full implicit answer; one cannot logically deduce an answer from the statements in the story without using general knowledge about the world such as:

(5) One can often get money from PCs.
(G) The hard part of getting money from a PR is getting It out. Once that Is done one can be said to have the money.
(7) Shaking helps set money out of a PB.

So In order to understand a children's story we need a theory of every day knowledge. This theory would have to answer questions like "What is the knowledge we have?" and "How Is It organized so we can get at the necessary Information when It Is needed?" Note that this latter quest Ion assumes that we have some specific task or tasks In mind, in our case answering questions about children's stories.

The rest of this paper divides into two parts. In the first part a rough description of a model of children's story comprehension will be presented. In the second section we will assume the model presented In the first and look at some narrow questions concerning the organization and content of our knowledge about piggy banks,

## 2 A Model of Children's Story Comprehension,

The model presented here Is solely concerned with deduction and does not consider problems of natural language per se. In particular It does not deal with syntax or those problems on the boundary between syntax and deduction like disambiguation of word senses and determination of noun phrase referents. (However, my Ph.D. thesis considers the noun phrase problem in some detail.)

So we will assume that as the story comes into the program It is immediately translated Into an internal representation which Is convenient for doing deduction. The internal representation will be a group of "assertions" each assertion being a predicate on an arbitrary number of arguments. Putting an assertion into the data base Is to "assert" it. The model will try to "fill in the blanks" of the story on a line by line basis. That Is, as It goes along. It will try to make connections between events in the story (usually causal connect ions) and fill In missing facts which seem important such as Janet's now having the money in (1).

### 2.1 Demons and Base Routines

Consider a fact like:

(8) If "It (s Cor will be) raining" and
If "person P Is outside"
then *'P wi 11 get wet"

We have an Intuitive belief that (8) Is a fact about "rain", rather than, say, a fact about "outside." Many things *happen* outside and getting wet Is only one of them. On the other hand only a limited number of things happen when It rains.

We will enbody this belief In our system by associating (8) with "rain" so that only when "rain" comes UP In the story will we even consider using rule (8). We will say that rain Is the "topic, concept" of (8). To put this another way, when a concept Is brought up In a story, the facts associated with it are "made, available" for use In making deductions. (We will also say that the facts *are "out in*" or "assert.") So, if "circus", say, has never come up, the program will not be able to make deductions using those facts associated only with "circus."

Mote however that we are not saying that "rain" has to be mentioned explicitly in the story before we can use (8). It is only necessary that there be a "rain" assertion put into the data base. Other parts of the story may provide facts which cause the program to assert that it is raining. For example:

body (an arbitrary program). We will execute the body of the fact only when an assertion Is In the data base which matches the pattern. (We will also say that the assertion "excites" the fact.) In (8) the pattern would be "someone outside." Then In (11) when we introduce (8) no assertion matches the pattern. But the next line creates a matching assertion, so the fact will be excited. We will say that a fact is "looking forward" when Its topic concept appears before the assertion which matches the pattern. When the assertion which matches the pattern comes first we will say that the fact Is "looking backward" (as in 10).

We can see how Important looking forward Is with a few examples.

(12) "Janet was thinking of getting Jack a ball for his birthday. When she told Penny, Penny said, 'Don't do that. Jack has a ball.'" Here we Interpreted the line "Jack has a ball" as meaning that he did not want another. The common sense knowledge Is the fact that In many cases having an X means that one will not want another X. This piece of information would probably be filed under "things to consider when about to get something for somebody else." Naturally It was an earlier line which mentioned that Janet was thinking of getting Jack a ball.

(13) "Bill offered to trade his pocket knife for Jack's dog Tip. Jack said ') will ask Janet, Tip Is her dog too.'" The last line Is Interpreted as the reason Jack will ask Janet. This requires information about the relation between trading and ownership.

(14) "Janet wanted to get some money. She found her piggy bank and started to shake It. She didn't hear anything." The last line means that there was nothing In the piggybank on the basis of facts about plggybanks.

In each of these cases it Is an earlier line which contains the Information which is used to assign the interpretation. So in (12) there is nothing Inherent In the line "Jack has a ball" which means "don't get him another." If there were, something In the line would also have to key a check for the following situations:

(15) Bill and Dick wanted to play baseball. When Jack came by Bill said "There Is Jack. He has a ball."

(16) Tom asked his father If he would buy him a ball. "Jack has a ball," said Tom.

(17) Bill's ball of string was stuck in the tree. He asked Jane how he could set it out. Jane said "You should hit It with something. Here comes Jack. Me has a ball."

Those familiar with Planner might notice that our "facts" look quite simillar to Planner antecedent theorems, with the exception that our facts can "look hack" as well as "look forward." Antecedent theorems are only designed to look forward. I Initially formulated facts as antecedent theorems because 1 was so Impressed with the need to "look forward." However, rather then call the facts antecedent theorems, I call them "demons" since It is a shorter name.

Specification and Removal of Demons. It should be emphasized that the model does not "learn" the informatlon contained in the demons. This information is put In by the model maker. On the other hand, the demons are not specific to the story in the sense that they mention Jack, or "the red ball." Rather, they talk about "a person X" who at one point in the story could he Jack, at another. Bill. We will assume a mechanism for binding some of the variables of the demon ("spicifying" the demon) at the time the demon Is asserted.

We want demons to be active only while they are relevant to the story, A story nay start by talking about getting a present for Jack, but ultimately revolve around the games played at his party. We will *need* some way to remove the "present getting" demons when they have outlived their usefulness. (An irrelevant but active demon not only wastes time and space, but can cause us to misinterpret a new line.) As a first approximation we will assume that a demon is declared Irrelevant after a given numer of lines have gone by.

Base Rout ines. So far we have said that demons are asserted when the proper concept has been mentioned. But this Implies that there is something attached to the concept name telling us what demons should be put in,

if we look at a particular example, say (13), it is Bill's offer to trade which sets up the context for the rest of the fragment. I will assume that the Information to do so is in the form of a program. Such routines, which are available to set up demons, will be called "base routines."

These base routines will he responsible for more than setting up demons. Suppose we are told that Jack had a ball, and Bill a top. Then Jack traded his ball to Bill for the top. One question we might ask is "Who now has the top?" Naturally since questions of "who has what" are Important in understanding stories we will want to keep tabs on such Information. in this particular case, it must again be the "trade" statement which tells us to switch possession of the objects. Every time a trade occurs we will want to exchange objects, so whenever we see "trade" we execute the "trade" base routine. Of course, the program can't be too simple-minded, since It must also handle "I will trade..." and perhaps even "Will you trade ...?"

A good test as to whether a given fact should be part of a base routine or a demon is whether we need several lines to set it up or whether we can Illustrate the fact by presenting a single line. (naturally several lines could be made Into one by putting "and's" between them, but this Is dodging the point. I am only suggesting an Intuitive test.) So we saw that "Jack has a ball" was not enough by Itself to tell us that Jack does not want another ball. Hence this relation Is embodied by a demon, not a base routine. On the other hand, often a single line can tell us quite a bit as In "Jack was on second base." This indicates that the base routine for "second base" can often tell us that we are talking about a baseball game.

## 2.2 BooKKeeping and Fact Finders

Undating. and. Bookkeeping. Up to this point we have Introduced two parts of the model, demons and base routines. In this section we will introduce the remaining two parts.

Again let us consider the situation when Jack had a ball, Bill a top, and they traded. When we say that Bill now has the ball. It implies that Jack no longer does. That is to say, we must somehow remove the fact that Jack has the ball from the data base. Actually we don't want to remove it, since we may be asked "Who had the ball before Pill did." Instead, we want to mark the assertion in some way to indicate that It has been updated. We will assume that there is a separate section, pretty much Independent of the rest of the model, which is responsible for doing such updating. We will call this section "bookkeeping."

Fact Finders. But even deciding that one statement updates another requires special knowledge. Suppose we have;

(18) Jack was In the house. Sometime later he was at the store.

If we ask "Is Jack In the house?" we want to answer "No, he Is at the store." Cut how is bookkeeping going to figure this out? There is a simple rule which says that (<state> A B) updates (<state> A C) where C Is not the same as B. So (AT JACK FARM) would update (AT JACK NEW-YORK). But in (18) we can't simply look for Jack AT <someplace which is not the store>, since he is in the house. To make things even worse, we could have:

(19) Jack was in the house. Sometime later he was In the kitchen.

To solve this problem we will need:

(20) To establish that PERSON is not at location LOC.

```
Find out where PERSON is, call it X.
If X = LOC , then theorem is false
so return "No."
If X is part of LOC then return
"No."
If LOC is part of X, then try to
find a different X.
Else return "Yes."
```

In (18) the bookkeeper would try to prove that Jack is not at the store, and it would succeed by using (20) and the statement that Jack is in the house. Bookkeeper would then mark the earlier statement as updated. Theorems like (20) are called "*fact finders*."

Like demons, fact finders have a pattern and a body. A particular fact finder is called when either a demon, base routine or bookkeeping wants to establish a goal which matches the fact finder's pattern. This is different from demons which are called when we encounter a given fact.

The basic idea behind fact finders is that they are used to establish facts which are comparatively unimportant, so that we do not want to assert them and hence have them in the data base. So in (18) we do not want to assert "Jack is not in the house" as well as "Jack is at the store." In the same way we will have a fact finder which is able to derive "<person> knows <fact>" by asking such questions as "was the <person> there when <fact> was mentioned or took place?" Again, this information is easily derivable, and not all that important, so there would seem to be no reason to include it explicitly in the data base.

## 3 Some Narrow Questions

In section 1 we stated that our theory of knowledge should answer questions like "how do we access the information" and "what is the information." In this portion of the paper we will look at two problems, one of each kind. We start with a question of information access.

### 3.1 Demon-Demon Interaction

In the description of the model it was stated that demons are excited when an assertion enters the data base which matches the demon's pattern. In this section we will present evidence that, given the model of part 1, we must also allow demons to excite other demons. I call this "*demon-demon interaction*."

*A Demon About PBs*. Before we can talk about demon-demon interaction we need to establish the need for two particular demons. Suppose we were given:

(21) Janet needed money. She got her piggy bank.
(22) Janet got her PB. "I want a nickel" she said.

Were we asked what Janet is going to do with the PB we would say, "get money from it." The obvious way to handle this is with a demon which is declared relevant when we see a person getting a PB. Naturally this would be done by the PB base routine. This demon would be looking for "<getter> need <money>" (i.e., that would be its pattern). When excited the demon would assert that there is a causal relation between "need money" and "get PB." We will call this demon PB-NEED-MONEY.

Now we might claim that "want money" should put in a demon to look for "get PB" rather than vice versa. However, this seems to be less reasonable, since there are many ways of getting money, but only a very limited number of reasons a person gets a PB.

*A Demon About Buying*. Buying things often requires money. So if we saw

(23) Janet was going to by some candy. She needed some money.
(24) Janet needed some money. She was going to buy some candy.

we would assert that the reason she needs money is to buy candy. Since the "need money" statement can occur on either side of the "buy" statement, it is clear that we want "buy" to put in a demon which says "If the 'buyer' needs money, it is because of the 'buying'." We will call this the BUY-NEED-MONEY demon.

We do not want "need money" to put in a demon looking for "buy." There are many things one can do with money, bribe a juror, pay rent, take a taxi, tip a bellboy, etc. It would seem more obvious to have the various events state that they (usually) require money, rather than have "need money" look for all of them. Nor does it seem reasonable to claim that all of these events ("bribing", etc.) are really versions of "buying." To express "take a taxi" as "buy some of the taxi driver's time plus the temporary use of the automobile in order to convey onself from one loaction to another" seems somewhat forced. Most of this "definition" comes from knowledge of economics rather than taxis.

*Evidence of the Phenomenon*. But now consider the following fragment:

(25) Janet was going to buy some candy. She went to get her PB.

Here we want to assert that Janet gets her PB because she wants money. But this time there is no previous assertion which says that she needs money. Of course, what is at work here is the fact that buying requires money. But, we have represented this fact as a demon, and so far we have no way for two demons to interact with one another. Now both PB-NEED-MONEY and BUY-NEED-MONEY will have the same pattern. So we can account for (25) by allowing a demon to excite other demons with the same pattern. Naturally this is what I mean by demon-demon interaction.

*A Restriction on Demon-Demon Interaction*. Demon-demon interaction is probably more complex than we have indicated so far. Consider:

(26) Janet was going to buy some candy. She was also going to buy some fruit.

In (26) both occurences of "buy" will activate BUY-NEED-MONEY demons. (Though we did not comment on this earlier, the Idea of specifying demons as mentioned in 2.1 obviously requires separate copies of a demon to be able to exist simultaneously.) However, (26) does not Imply that Janet really needs money. For all we know she has as much as she needs In her pocket. If demon-demon Interaction were as simple as we have made it out to be, the two instances of BUY-NEED-MONEY would join up to produce a "need money" assertion. So It Is not sufficient for two demons to be looking for the same pattern.

Looking at example (25) we note that one of the demons cave a reason why Janet might need money, and the second suggested that needing money was the cause of a certain action. So we have:

Will buy --> Need money --> Will get PR

To put this in everyday terms, in (25) we have both a motive for needing money (buying), and a result of needing the money (go and get PB). In (26) we have two motives. The natural suggestion is that demon-demon interaction be restricted to cases where we have both motive and result.

How do we recognize when we have both motive and result? As It stands now one demon looks pretty much like any other. We might just try to label all demons as "motive" or "result" with respect to their pattern. On the other hand It might be possible to derive "motive" and "result" from more basic considerations. At any rate, it seems premature to formalize such concepts at this point. We simply don't know enough.

Capturein Generalizations Before moving on I should point out that the kind of argument used in this section (and the next also) Is a "capture the generalization" type argument commonly found In linguistics. We could have created a new demon to explain (25). It would have said, "If a person gets his PB look for him planning to buy something." However, this would be missing the generalization that "motives" and "results" always act this v/ay. So far I have only given one example to' support the "demon-demon interaction generalization", but in In the next section we will see another.

### 3.2 Putting Money into a Piggy Bank

In this section we will look at a possible demon associated with piggy banks and argue that the deduction ft would account for can be better handled by demon-demon interaction between two other demons. In effect we will be trying to determine, on an extremely small scale, what people know.

A PIGGY Bank Problem. One fact we know about Pg's is that they are good places to keep money. This fact seems to come Into play In:

(27) Penny said to Janet, "Don't take your money with you to the park. (You will lose It.) Go and get your PB!"

(28) After Janet helped Ms. Jones with her groceries Ms. Jones gave her a dime. Jack came along and said. "Cone with me to the park, Janet." "OK," said Janet. "Hut first I an going home to find my PB, I do not want to take the money to the park."

(29) Janet put some money on the sink, Mother said, "If you leave the money there it may fall In the drain. Let's find your PB."

In each case the natural question is, "Why should Janet get her PB?" Now we might try to construct a "piggy bank" demon which responds to some common element in (27) - (29) and then make the necessary assertions. A close look at the examples even gives a start at what such a common element might be. say "a particular location for the money is negatively evaluated." We will call this demon PB-BAD-PLACE. The trouble with such a solution would be that It would not account for:

(30) Janet said, "I am going to put my money away. I will get my PB."

(31) Janet helped Ms. Jones with her groceries. Ms. Jones gave Janet a dime. Jack came along and said, "Janet, let's go to the park." "OK," said Janet. "But i want to put my money In a safe place. I am going to get my PC."

flow there is nothing saying that our demon needs to account for (30) and (31). However, it seems quite obvious that we are using the same information In ail the examples above, The only difference Is that In (27) - (29) we are expressing the need for a "safe place" by making negative comments about another location. If this Is a single fact we would like a single demon to express It. The trouble is finding what (27) - (31) have In common.

A Non-PJfrsv Bank Problem. in the course of looking at examples like (27) - (31) 1 noted examples like;

(32) Penny said to Janet, "Don't take your money with you to the park. Put It on the shelf."

(33) After Janet helped Ms. Jones with her groceries Ms. Jones gave her a dime. Jack came along and said "Come with me to the park, Janet." "OK," said Janet. "But first I am going to put my money in the house, I do not want to take the money to the park."

(34) Janet put some money on the sink. Mother said, "If you leave the money there It may fall In the drain." Janet put the money In a drawer.

(35) Janet said "I am going to put my money away. I will put it in my toy box."

(36) Janet helped Ms. Jones with her groceries. Ms. Jones gave Janet a dime. Jack came along and said "Janet, let's go to the park." "OK," said Janet. "But I want to put my money In a safe place," Then Janet went Into' the house and put the money In her room.

These examples exactly mirror (27) - (31), except that (32) - (3E) don't mention PB's. Naturally, In these examples the question to ask Is "Why did Janet put the money In the drawer?", "In the house?", etc.

Such examples tend to Indicate that the problem facing us Is wider that just PB's. We will name this wider problem the "put away" problem. However It Is not the case that our problem with PB's can be completely reduced to the "put away" problem. So while In the non-piggy bank examples wo mention that Janet has or actually Intends to "put" the money some place. In the PB examples all we needed to say was that Janet was going to get the PB. To put this another way, our knowledge of PC's allowed us to Interpret "get PB" as meaning that Janet was going to put money into it. However our knowledge of houses or shelves does not allow us to make similar deductions In (32) - (36).

The put-Away Demon. Ignoring piggy banks for the moment, what would a solution to (32) - (36) look like? We will have some demon, cal1ed the PUT-AWAY demon, which Is activated by lines like:

(37) Don't leave the money by the sink.
(38) 1 do not want to take my money to the park.
(39) I will put my money away.

These lines will put In a demon looking for "put away" and the demon will assert that the reason for putting the thing away Is (37) - (39). Ultimately we will want a theory of why people put things away (i.e., what lines put In the "put away" demon), and how to determine what constitutes a "put away" 1ocat ton. However, (32) - (36) clearly show that the problem Is distinct from the question of what we know about PBs.

The Plggv Bank Demon. Whet we will now see is that if we assume the PUT-Away demon, all the examples In (27) - (31) fall out easily, plus a few others which we haven't even looked at yet. But first we need to consider a new PB demon entitled PB-MONEY-IN. It Is parallel to PB-NEED-MONEY, but while the latter was for recognizing that money was going to be taken out of he PB, PB-MONFY-IN Is for recognizing that money Is going to be put In. It says "If you see that the person wants some money to be In the PB then the reason he Is getting the PB Is to put It In." (Actually this theorem Is true of a wide class of containers, but that does not affect the argument at hand,) This demon will account for examples tike:

(40) Ms. Jones gave Janet a dime. Janet went to get her PB. "I want the money to be In my PB," she thought.

(41) Janet got her PB and dropped some money In.

*(42)* After Ms Jones gave Janet a dime, Jack came by and asked Janet If she wanted to go to the park. "OK," said Janet. "I will go home first and get my PB." Soon Janet came back and said "My money is In the PB, let's go!"

Demo-Demon Interacion.- Now, If we assume demon-demon interaction as discussed In section 3.1, PB-MONEY-If plus PUT-AWAY will Interact to solve all the examples from (27) to (31). Let us see how this will happen.

First note that the restrictions we placed on demon-demon interactions are met here. First both demons have the same pattern, e.g., "money is in PB." (Actually the pattern for PUT-AWAY Is "<object> is In appropriate location)" however <object> will be bound to the money at the time the demon is asserted, and appropriate location) will match PB when the demon is excited.) Secondly, we need both a motive and a result before we can "combine" demons. In the case at hand, PUT-AWAY Is a motive for having the money In the PB, and "get PB" Is a result of intending to put money In the PB.

Saving Money. Finally, note that our solution extends to the following case:

(U3) Janet got a dime from Ms. Jones. She said "I am saving my money to buy a bicycle. I am going home to get my PB."

Here we know that Janet Is going to put the money In the PB because of the "save" statement. However, we Immedlately note that we have cases 1i ke:

C4l4) Janet got a dime from Ms. Jones. Janet told her "I am saving my money to buy a bicycle. I am going home to put the money away. (I am going home to put the money In my drawer.)"

Naturally, *(kk)* Indicates that "save" must activate PUT-AWAY. If this Is the case, then (i»3) Is accounted for in exactly the same manner as all the Initial examples. While the reader may not be surprised at this result. I am, since Initially I thought that the relationship of "save" with piggy banks would need a separate PB demon.

## It Conclusion

The two halves of this paper stand In contrast to each other. The presentation of the nodel (section 2) Is general (In theory covering all of children's stories), but vague and full of covert appeals to the reader's intuition. Section 3 on the other hand is narrow, only talking about small portions of our knowledge of PCs, but tightly reasoned (hopefully).

flow by themselves the conclusions of section 3 are not that Important. Of course. If we could pin down one hundred facts the way we pinned down one in section 3.2 then vie would have the beginnings of a theory of knowledge. Dut 1 did not write this paper to tell of one fact about PCs. Rather I view the paper as an Illustration of how one might go about the task of constructing a theory of knowledge.