"Shakespeare's Plays Weren't Written by Him,
But by Someone Else of the Same Name"

An Essay on Intensionality
and Frame-Based Knowledge Representation Systems

Douglas R. Hofstadter, Gray A. Clossman, and Marsha J. Meredith
Computer Science Department
Indiana University

July, 1980

TECHNICAL REPORT NO. 96

"SHAKESPEARE'S PLAYS WEREN'T WRITTEN BY HIM,
BUT BY SOMEONE ELSE OF THE SAME NAME"

DOUGLAS R. HOFSTADTER
GRAY A. CLOSSMAN
MARSHA J. MEREDITH

JULY, 1980

*Research reported herein was supported in part by
the National Science Foundation under grant number
NSFMCS 80-09201

ABSTRACT

By analyzing a large number of typical fragments from everyday conversations and a few vignettes from real life, we develop a set of distinctions between references to things and references to the roles they fill. We discuss these distinctions in the language of frames (slots, fillers, pointers, and so on). The process of referring to something by its name but meaning its role, and the converse, mentioning a role but meaning the specific thing that fills it, are shown to be closely related to the mechanisms of analogical thought -- particularly the production of "conceptual skeletons" and their use as very high-level "addresses". The sense of uniqueness or tangibility of a mental node, and some properties which are necessary for it, are explored through analysis of the title sentence and other examples. We conclude with a set of challenging anecdotes which explore the connection between roles filled by something, and our emotional sense of the unique identity, or "soul", of that thing.

Our purpose is to address certain philosophical questions in the spirit of implementors of a knowledge representation language, with the hopes that some light may be cast upon both knowledge representation and the philosophical issues themselves.

"Shakespeare's Plays Weren't Written by Him,
But by Someone Else of the Same Name"

An Essay on Intensionality
and Frame-Based Knowledge Representation Systems

FIVE FRIENDS

Consider the strange case of five friends, each of whom says, "I want to own the fastest car in the world" (a sentence from [Winograd 72]).

Tom: He wants always the CURRENTLY fastest car -- today a red Ferrari, tomorrow a green Mustang, Friday a purple VW... He doesn't care what kind it is, as long as the car he owns is listed in the record book.

Dick: He took a ride in the world's fastest car and loved the feel of it. He has forgotten what kind of car it was, even its color and everything else -- all he remembers is that it was the world's fastest car, and so that's his only way of describing the car he wants to own -- his only "handle" to that car.

Bill: He wants a certain red Ferrari which he has driven a lot and knows intimately, but which is hard to describe to most people. It happens, however, to be the fastest car in the world, so that's how he describes the object of his desire to his friends (although actually he doesn't give a hoot about that feature of the car).

Harry: A hot-rodder from way back, he wants his beloved green Mustang ("Old Mussie") to BECOME the fastest car in the world, and every day after work he puts in four hours souping it up.

Pete: He is green with envy of the fame and glory that presently accrue to Luigi Borsari, owner of that wonderfully swift red Ferrari. Luigi's got all that dough and those swell women... Pete wants to BE THE OWNER of the fastest car in the world!

The same sentence, "I want to own the fastest car in the world", validly expresses the desire of each of these men. How can we formally represent the differences between the five underlying meanings?

To solve this puzzle and to raise more difficult, but similar, puzzles is the purpose of the present paper. Our puzzles poke at the boundaries of the expressive ability of traditional frame-based knowledge representation systems in artificial intelligence research. All of our puzzles come from everyday sentences and experiences, and reveal how deep and flexible are many processes of ordinary thought.

Diagrammatic "solutions" are given for some of our puzzles, but in the balance we pose more questions than we answer. What we mean by a "solution" is not a complete frame-based representation system which can handle all the puzzles, but a set of clearly mapped-out distinctions which any implemented system (including one we intend to develop) would have to be able to make, and a notation for these distinctions. Before tackling our many puzzles, including those of the fastest car and the title sentence, we give a short historical overview of the concept of intensionality, which is of the essence here.

CLASSICAL EXAMPLES OF INTENSIONALITY

A naive sense of logic tells you that if two phrases denote the same thing, then the meaning of any sentence using one of the phrases should be unchanged if you replace it by the other. But consider the following sentence: "The Morning Star is visible only for a short period just before sunrise." It so happens that the so-called Morning Star is actually the planet Venus. It also happens that there is a so-called Evening Star which is likewise brilliant, and which, needless to say, is visible only for a short time just after sunset. This celestial object turns out to be the very same planet -- a fact of which most people are ignorant. Given this fact, then the naive premise would have it that the sentence's meaning stays the same under substitution: "The Evening Star is visible only for a short period just before sunrise." But clearly the two sentences mean very different things. The first is a familiar statement to anyone who has observed the Morning Star, while the second is paradoxical-sounding, unless you know its hidden resolution. These sentences DON'T have the same meaning. This is the classical example of INTENSIONAL versus EXTENSIONAL qualities of descriptive phrases (to be defined in a moment).

Another humorous classical example concerns George III of England. Once he asked, "Who is the author of 'Waverly'?" When told "Sir Walter Scott", he complained, "That's of no help! My question is, 'Who is Sir Walter Scott?'" (It is ironic that today, were one to ask a philosopher who George III was, the answer would likely be, "Some king who asked who the author of 'Waverly' was.")

A more contemporary example is provided by Lewis Creary [Creary 1979]. It runs this way. Suppose Mike has heard from Jim that Jim's wife is an avid birdwatcher, as he (Mike) is. So now Mike wants to meet Jim's wife. That's simple enough. However, unbeknownst to Mike, she happens also to be Sally's mother. So does Mike want to meet Sally's mother? Certainly HE wouldn't say so. He isn't aware of her in that capacity, in that role. He wants to meet a person who fills a different role -- that of Jim's wife. On the other hand, it's not BECAUSE she's Jim's wife that he wants to meet her, but because she's a birdwatcher. It's just that the role of "Jim's wife" is the only handle he has by which to "retrieve" her.

Clearly the problem in all these examples is, in what sorts of contexts can one "referring expression" be substituted for another without change of meaning? This is the issue of "referential transparency" versus "referential opacity" (Gottlob Frege's terms). It is intimately connected with the intensional-extensional distinction. So what is this distinction, exactly?

EXTENSIONS VERSUS INTENSIONS

Although the concepts of extensionality and intensionality in their essence are rooted as far back as medieval arguments over the meaning of generic and abstract terms such as "man", the actual words "extension" and "intension" originally came from predicate logic and set theory. Here, one can think of a set either in terms of its EXTENSION (the totality of its members) or in terms of its INTENSION (the predicate which its members all must satisfy). For example, the set INTENSIONALLY specified by the phrase "primes under 10" has as its EXTENSION the numbers 2, 3, 5, 7. One can define sets whose members one has a much harder time exhibiting -- for example, the set of all odd perfect numbers (a "perfect" number being one whose proper factors add up to it, such as 28 = 1 + 2 + 4 + 7 + 14). No odd one is known, yet there may be an infinity of them. Nonetheless, we can deal as easily with the intension of this mathematical concept as with the intension of any other mathematical concept.

The intensional-extensional distinction can be extended to apply to descriptive phrases and their denotations in computer languages or in human memory. By "descriptive phrase", we mean any expression or symbolic structure -- including a mere name -- that denotes (or seems to denote) an entity. Let us then turn to the case of interest to us -- knowledge representation systems in computers or people. It might seem natural to say that a descriptive phrase in a knowledge representation system is an INTENSION, whose EXTENSION is the thing which it denotes. Thus, the descriptive phrase "the current President of the United States" is an INTENSION whose EXTENSION is Jimmy Carter. One might even say, more picturesquely, that the phrase acts as a "pointer" that points to Jimmy Carter.

Actually, we have hit upon the crux of the problem right here. The word "pointer" is not an abstract philosophical term meaning "denotation", but a computer science term meaning, roughly, "address". A pointer is a data structure in memory -- you can think of it as an arrow both of whose ends are structures INSIDE THE COMPUTER MEMORY. So how can a pointer -- something INSIDE our knowledge representation system -- point to a person -- something OUTSIDE of the system? In our mathematical and set-theoretical examples, the things pointed at (numbers) and the things doing the pointing (expressions) are all parts of the same system. The phrases "3 factorial", "smallest perfect number", "5 + 1", and the number 6 all belong to number theory. The thing pointed at is not "outside the system".

Of course, things are not quite that simple. The question of the real existence of Platonic numbers "out there" is still much argued over. Many philosophers would maintain that even here, pointers point outside the system. Yet in some sense, the entities pointed to and the pointers themselves form a closely integrated conceptual network, a closed universe whose exploration is the purpose of mathematicians. To discover that Jim's wife (the ratio of circumference to diameter) is the same as Sally's mother (the sum of some infinite series), one stays completely within the confines of a well-defined conceptual territory, obeying certain kinds of rules.

But getting back to Jimmy Carter, we see that the whole problem is that we don't have PEOPLE in memory, but rather, SYMBOLS. We can't have a pointer to a PERSON -- only a pointer to some sort of memory structure that is supposed to REPRESENT a person. This may not seem so tough -- just make one "node" (a coherently organized collection of information in memory, characterized by being addressable as a unit) for each person you know or know of. Let's see, there's one node for Jimmy Carter, one node for Richard Nixon, one for George Washington, one for Jim's wife, one for Sally's mother -- whoops! We have TWO nodes for one person here. What's going on? As long as we stuck with NAMES, we seemed to have exactly one node per person -- but then we switched over to descriptive phrases -- intensions -- and the neat one-to-one match couldn't be maintained.

MERGING NODES

It is inevitable in any representational formalism that distinct nodes will sometimes be manufactured that, it turns out later, stand for the same thing. One has incomplete information on some entity -- two or more orthogonal "perspectives" -- and one fails to realize that both views are of one object. What happens when one finally finds out that the two nodes represent the same thing? Clearly, they have to be fused somehow into one new node. (Some people would prefer to speak of growth of links between separate nodes, rather than of fusion of the nodes themselves. In either case, we have no idea how this would be done in the brain itself, since we have at present no knowledge of of the relations between the symbolic level and the physical level in the brain. Despite our current state of ignorance on the actual "implementation" in the brain, there is indirect experimental evidence [Anderson 1978] suggesting that the nodes remain distinct, and that links grow between them.) In the case of Jim's wife and Sally's mother, it seems quite straightforward. Since there are no overlapping pieces of information, merging the two nodes simply amounts to creating a new node with all the features of the old nodes.

It gets trickier if the two nodes to be merged have potentially conflicting or overlapping features. For example, let us cook up a more complex version of our simple situation. Let's say Mike has a daughter named Peggy, who's in the fourth grade. Last week, Peggy was telling her parents about her new friend Sally. Peggy said, "Sally's mom is the best cook I ever SAW! And next Wednesday, you know what? She's gonna fix a strawberry shortcake for Sally's dad because it's his BIRTHDAY and they're gonna have a surprise BIRTHDAY party for him! How come WE don't ever have surprise birthday parties, huh?" On Monday at work, Jim tells Mike, "My daughter caught a cold last weekend while birdwatching with Maude." On Tuesday at dinner, Mike asks Peggy if she saw Sally at school. "Oh, no, Daddy -- she has the flu," answers Peggy. "That's too bad," replies Mike. Late the next afternoon, Jim's wife Maude stops by work specially to meet Mike. After a few minutes' pleasant birdwatching conversation, she mentions she has to go to buy some strawberries for a shortcake. Some bell rings in Mike's brain. What happens? We will come back to this exciting story momentarily...

ONE NODE FOR GEORGE WASHINGTON?

It is obvious from examples like this that information about, say, individual people is often scattered about in separate little packets which are distinct nodes that some trigger or other can cause to be merged. Ideally, of course, one wants ONE NODE PER INDIVIDUAL, and each such node should be indexed, most probably, by the NAME. Does this mean that NAMES should serve as the equivalent of EXTENSIONS? Some epistemologists maintain that this is, in essence, the case. But that leads one into many dilemmas, the title sentence of this paper being an example which we shall consider below.

Despite its impossibility, what would this ideal representation be like? If we were trying to maintain a strict one-to-one correspondence between nodes in our heads and, for instance, real people "out there", what are the characteristics that we would want, say, the "George Washington" node to possess?

The word "node" conjures up an image of a central point, a hub around which are clustered many pieces of information. The following facts about George Washington ought to be clustered around the GW hub:

(1) Name is "George Washington"
(2) Birthdate is February 22
(3) Was husband of Martha
(4) Was first President of the United States

Are some more important than others? Are some more central, more vital to the George-Washingtonhood of this node? If one or more were eliminated, would the node lose its George-Washingtonhood? Where does the essence of this node, its "Core ID" reside?

NAMES AND NODES

People tend to treat the NAME as the essence of a node. (In fact, just a few lines back, I unwittingly revealed such an attitude, by talking about possibly losing the "George-Washingtonhood" of the node. Why not worry about losing the "first-presidenthood"?) Since no pointer in a human memory (or a computer memory) can possibly point to the real person who was George Washington (or first President), we have to find a substitute, and the node with the NAME "George Washington" seems very natural. His name is by far the most common way we have of getting ourselves or others to retrieve the GW node. And yet, there are a lot of problems with thinking of that as the complete answer as to the identity of the node.

What if we find out that the first President had another name? What if we find out that the person named George Washington was born on March 1? What if we find out that the person born on February 22 was married to someone named Marsha? Just how many distinct people are there actually here, anyway?

This begins to sound like one of those amusing logic puzzles where you are given a host of sentences such as "The Norwegian's best friend wears red pajamas" and "The sailor's wife murdered the smoker" and by using all these fragmental descriptions of people, you gradually piece together a composite description of each person. It could turn out, for instance, that the Norwegian and the smoker are one and the same person. But you could just as easily imagine that the sailor's wife and the smoker are the same person.

In a game, it can be fun to take apart descriptions of people and put them back together in random combinations, but in real life it is disorienting. In the extreme case, any descriptive fragment could be taken as describing a separate entity. But if (as with George Washington), we attach them all to one node, we are showing that we feel there is a unique object "out there" which has all these properties. Yet potentially, each fact could be ripped out and attached to another node. Or conversely, new facts can always be added to this node. This comes, quite simply, from the fact that a node is not a person, and no amount of information on a node can ever equal the sum total of facts about a person.

Strikingly, though, we still tend to feel that descriptive phrases such as "first President" refer only INDIRECTLY to a person, whereas the NAME "George Washington" is somehow a "direct line" to the person. Thus we tend to visualize the memory equivalent of the PHRASE "first President" as being a POINTER to some node, whereas the memory equivalent of the NAME would be the NODE pointed to! But this is too simplistic. What about the man who works in the photography store, whom you've talked with dozens of times, yet have no name for? What about someone (I'm thinking of Cassius Clay) who has a religious experience and changes his name with great fervor (I'm thinking of Mohammed Ali) and insists his old name is completely irrelevant? You might have thought of Richard Alpert -- I mean Baba Ram Dass. What about the case where you repeatedly retrieve from memory the wrong name for someone you know well? In many families, a mother will call out her son's name when she means to call her husband, or vice versa. Certainly she doesn't confuse the PEOPLE, yet the names get crossed. (It is interesting that very seldom would she use her DAUGHTER's name when she means to call her husband. Why should this be the case?)

And what about things other than people? What about that red Ferrari which has no name but which is best described as "the fastest car in the world"? What about that splotch of paint near the light on the ceiling? Sometimes the best way to pinpoint something is not by name at all, but by careful description. Which is more useful -- to say, "Look at Arcturus!" or "Look at that star about ten degrees above the top of that poplar tree over there!"?

A name is just one way to address a node. Once you realize that every referring phrase (including names) should be thought of as a POINTER, then you begin to wonder what a node is. Things get blurry, because now ALL representations seem to be intensional. Is there such a thing as an extensional description or node? What gives a node its uniqueness?

FRAMES AND SLOTS

Let us now attempt to examine these issues in the terminology of frames. In this paper, a SLOT will be a pair: a NAME and a FILLER (or VALUE). As in the original and most subsequent frame literature (see especially [Minsky 1975]), a FRAME will be a meaningful collection of slots, with an addressable name -- in other words, a frame is the computational equivalent of a node. We will allow frames to overlap and to have as little as one slot. A LABEL is written at the top left corner of each frame. Sometimes the label is simply a proper noun, such as "Harry". But more often, the label consists of two parts: a word or atom (such as "Great-Author") and an arbitrary number, which is attached to make the label unique (although sometimes we leave out the number, for simplicity's sake). The atom indicates a CLASS that the frame belongs to, and from which some or all of its slot-names are inherited.

The FILLER of a slot may be either an atomic object (i.e., indivisible, not interpretable further within the frame system), another frame, or another slot. If the value is ATOMIC, it will be written to the right of the slot name. If it is a FRAME, it may be denoted, in a diagram, in either of two ways: (1) by drawing an arrow from the slot name to the value-frame, or (2) by writing the LABEL of the value-frame to the right of the slot name (this abbreviated notation makes diagrams cleaner). Either notation should be understood as representing a pointer in the frame data structure. Finally, if the value is a SLOT, a pointer will be drawn to it -- but this will require some detailed explanation below. Some pointers may have extra structures -- for example, backlinks or dates. It is important for "Name" pointers to be stored in both directions, so that nodes can be retrieved by name without any search.

POINTERS FROM SLOTS TO SLOTS

In Figure 1 we have shown possible representations for three meanings of our anomalous first sentence -- those of Tom, Dick, and Bill. Note that the difference is completely represented by the position of the tip of the pointer that points to the "Fastest" slot in the frame for the convenient (but probably nonexistent) document called "1980 Car Records Book". (We'll come back to the parenthetical point later.) In Tom's case, the pointer's tip points to the SLOT itself; in Dick's case, it points to the LEFT of the COLON, and in Bill's case, to the RIGHT of the COLON. Notice that what Dick and Bill want is a specific physical car, which for one reason or other they DESCRIBE as "fastest", whereas Tom has no preference as to the embodiment of his desire; he desires simply a ROLE. Here is a summary of the distinctions:

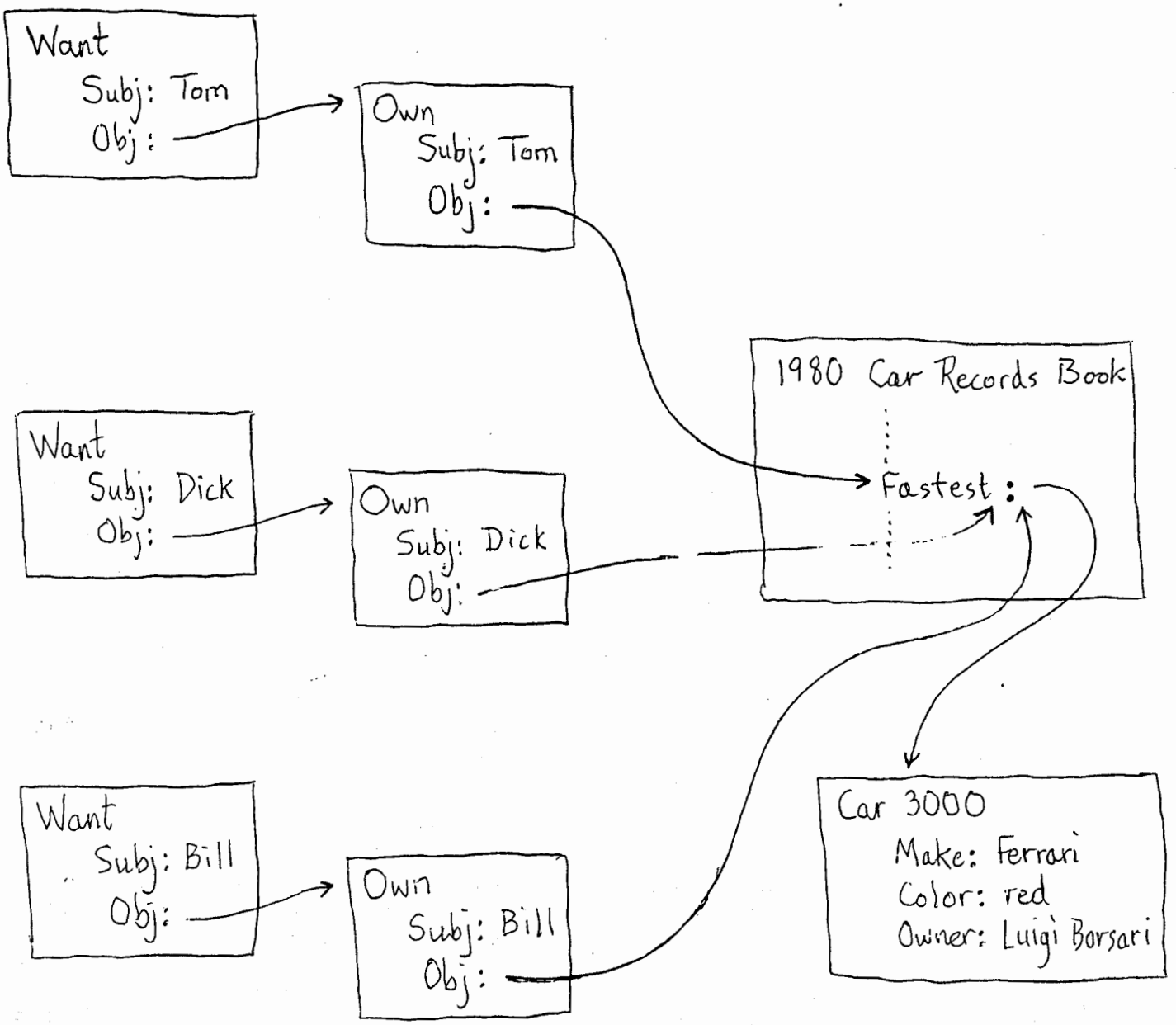


Figure 1. Three meanings for "I want to own the fastest car in the world."

Arrow pointing to SLOT NAME:
One is interested only in the ROLE and not in who plays it.

Arrow pointing to LEFT OF COLON:
There is a SPECIFIC OBJECT being pointed to here, but the speaker is addressing it through this role by UNFORTUNATE NECESSITY: he would prefer to be able to point at it directly, but cannot; this role is the only uniquely identifying "handle" he has for the object.

Arrow pointing to RIGHT OF COLON:
There is a SPECIFIC OBJECT being pointed to here, one which the speaker is familiar with, but for one reason or another he PREFERS to characterize it indirectly, by citing this particular role which it happens to play; it would be unhelpful, confusing, too much trouble, or possibly even embarrassing to name it directly.

The three distinctions can be applied to the phrase "Jim's wife". If Jim goes to the bank to get a loan and the officer says, "I'll need your wife's signature", here is a clear-cut case where what's important is the ROLE, not the person behind the description. The officer doesn't care if the next day Jim goes out and marries somebody else, he just wants the signature of whoever fills that slot. Mike is in the second situation -- he wants to meet this birdwatching nut about whom all he knows is that she is married to Jim. Finally, he meets her and gets to know her pretty well, and tells a friend, "Jim's wife and I often talk about birds." This is an example of the third case: he could have called her by name but CHOSE not to because the friend wouldn't know who "Maude" refers to.

JIM'S WIFE AND SALLY'S MOTHER: THE RECURSIVE INTENSIONAL MERGE

Speaking of Maude, let's return to the story of how Mike came to merge his nodes for Sally's mother and Jim's wife. To aid you in thinking about how merging of this sort might proceed, we've shown in Figure 2 a prototypical case of the "recursive intensional merge".

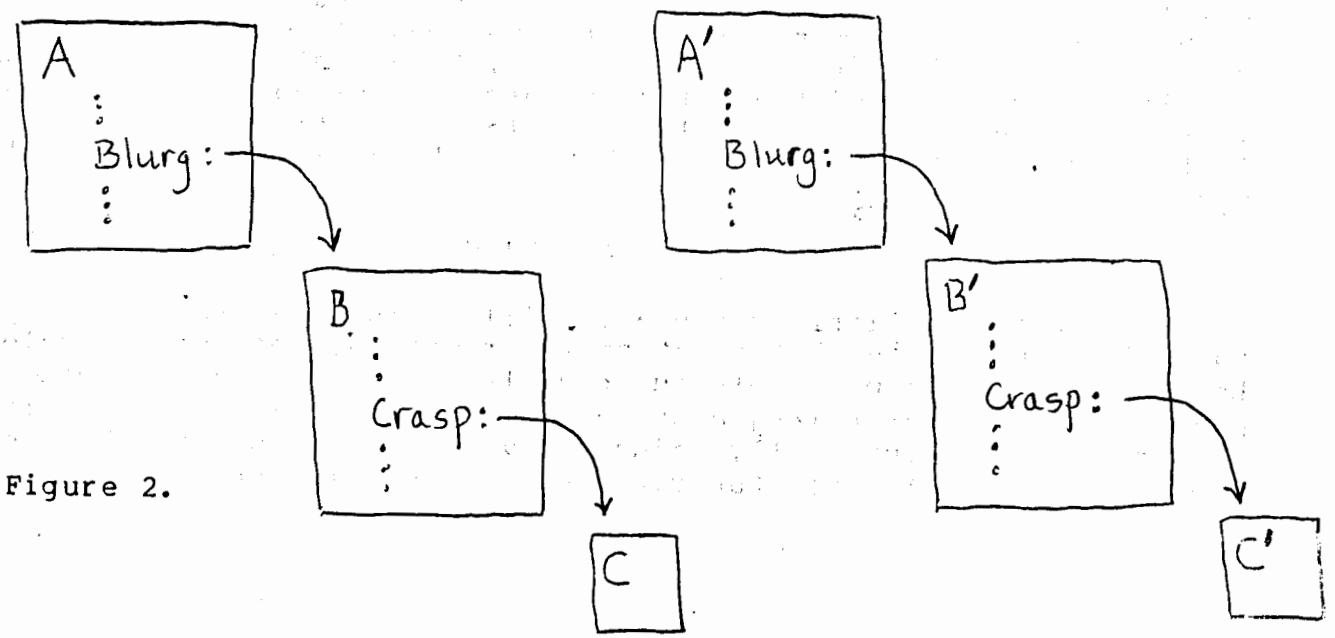


Figure 2.

There are two frames, A and A', each having a "Blurg" slot. If we find out that A and A' represent the same individual, then it follows that their blurgs are the same, so B and B' must also be merged! But now, B and B' both have "Crasp" slots, filled respectively by C and C'. Obviously, C and C' must now be merged also. The merging process cascades from one level to another until, finally, at some level, the frames involved have only a SIMPLE merge -- i.e., they have no apparently conflicting slot fillers.

What about the inverse case -- where we find out that C and C' should be merged? Are we then forced to merge B and B'? Only if the "inverse crasp" relation is unique -- i.e., only ONE object can have an object X as its crasp. If that is the case, then, of course, the recursive intensional merge proceeds UPWARDS. If "inverse blurg" is likewise unique, then we're forced to merge A and A' as well. Things can get really messy when both upwards and downwards types of recursion are combined in a single process. And that, of course, is what we have in our story.

In Figure 3, you see diagrams showing the state of Mike's knowledge before any merging began. It all begins with Mike's hunch that SSC-1 and SSC-2 might be the same strawberry shortcake. This induces (upwards) a merge of the two "Bake" frames (since you can bake a cake only once). A "Bake" has only one baker, so that forces the merge of the frames for Maude and Woman-1. Now we have some DOWNWARDS-induced merging, since both Maude and Woman-1 have "Daughter" and "Husband" slots. Let's merge the daughters first. Peggy's friend Sally is now seen to be Maude's daughter. But -- Maude is Jim's wife. Aha! Jim's wife is Sally's mother!. (Pi is the sum of that wonderful infinite series...) And of course, in merging Man-1 with Jim, Mike finally comes to the conclusion that today is ... Jim's birthday!

What about the merging of the diseases? We seem to have a conflict here. Peggy said Sally had the flu, while Jim said his daughter had a cold. Uh-oh -- does this invalidate everything? No, of course not. Mike knows to give little Peggy some leeway -- how would she know exactly what was wrong with her friend? More precisely, Mike backs off from his belief "Sally has the flu" and remembers its source. It turns into something more like "Sally has WHAT PEGGY CALLED the flu" -- a node for an ill-defined malady. This, then, matches the cold more easily. Such accommodation of sloppiness is, of course, so commonplace in human thought that it hardly calls for commenting on; the trouble is that it is not at all obvious how to make an artificial intelligence system that can do it.

POINTERS TO THE LEFT AND RIGHT OF A COLON

You may be a little confused, still, about the arrows that point to a slot, and those that point to the right and left of the colon. Let's consider one more colorful example to try to clarify this subtlety. You make a long distance call to Area Code 807, and this operator with a wonderfully seductive voice comes on. Oh, how you would like to meet her! But you hang up and the little romance is

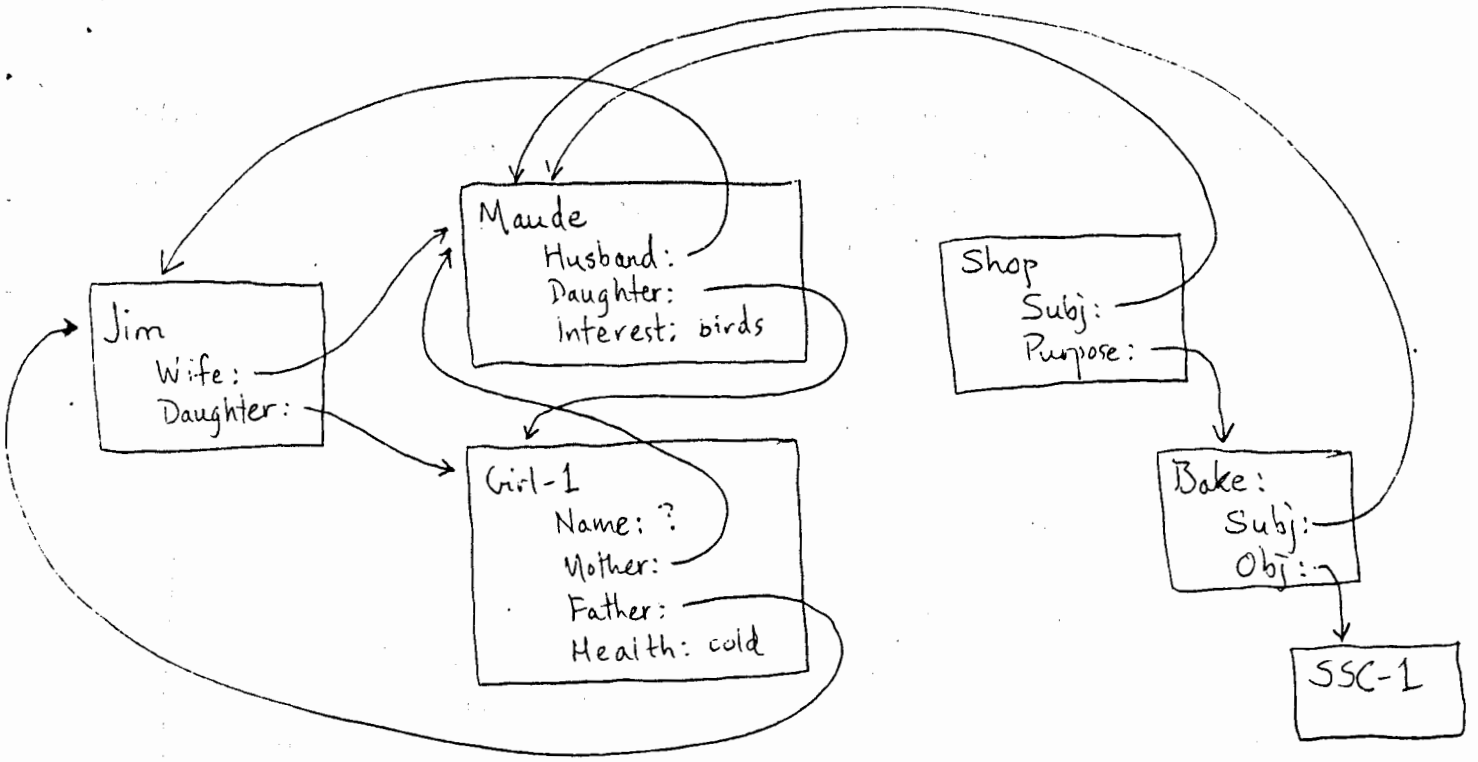
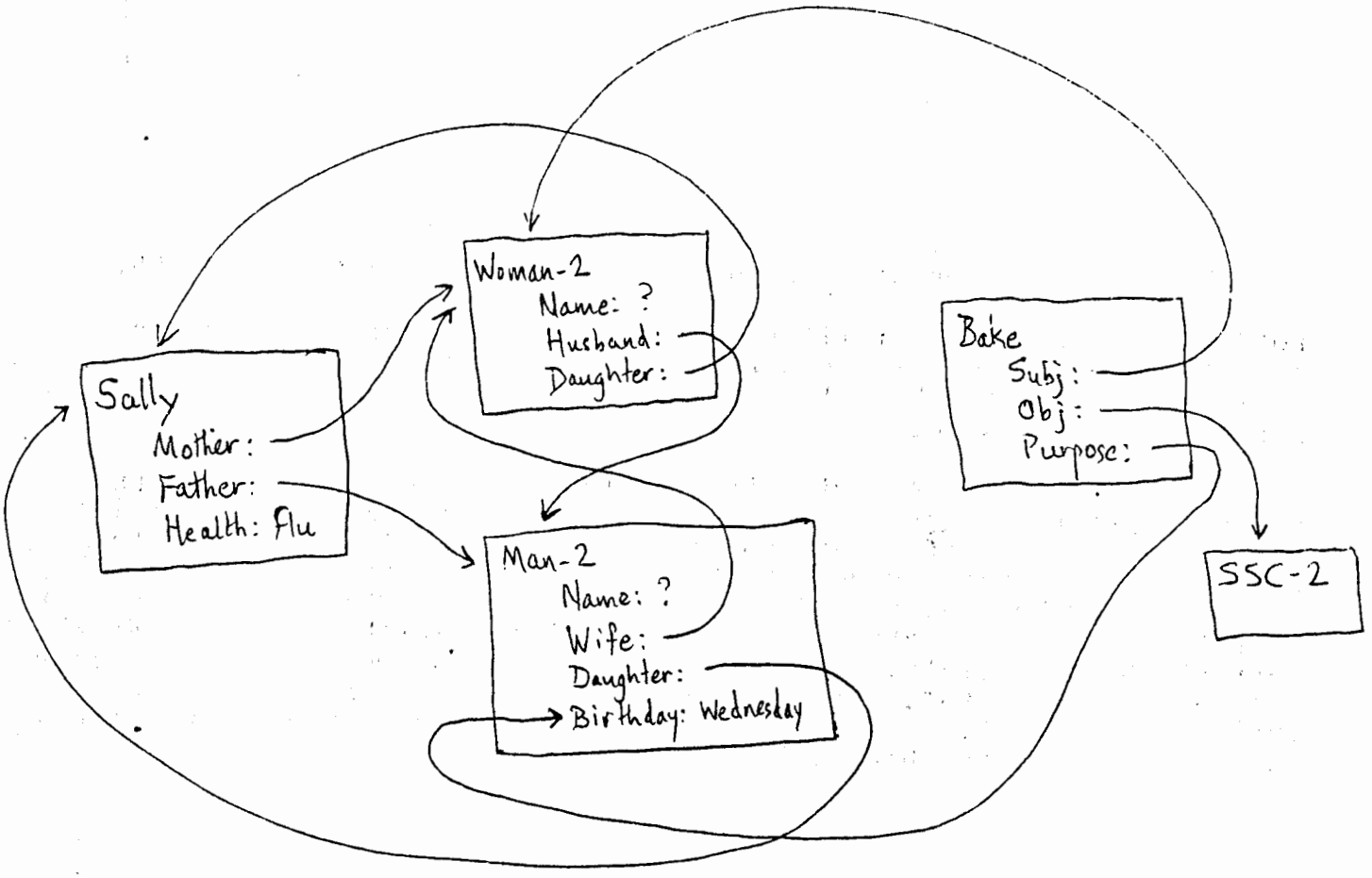


Figure 3. Mike's unmerged knowledge. Above, what Mike knows of Jim and his family. Below, what Mike knows of Peggy's friend Sally and her family.



over. How do you tell your friend about this woman? You have only one way of DESCRIBING her -- as "the operator when I called so-&-so in Area Code 807" -- yet it's not at all because she fills that ROLE that you want to meet her, it's because of her VOICE! Let's say you even know she's Operator 55 in that Area, so you have a unique tag that identifies her. Look now at Figure 4a. An arrow points to the

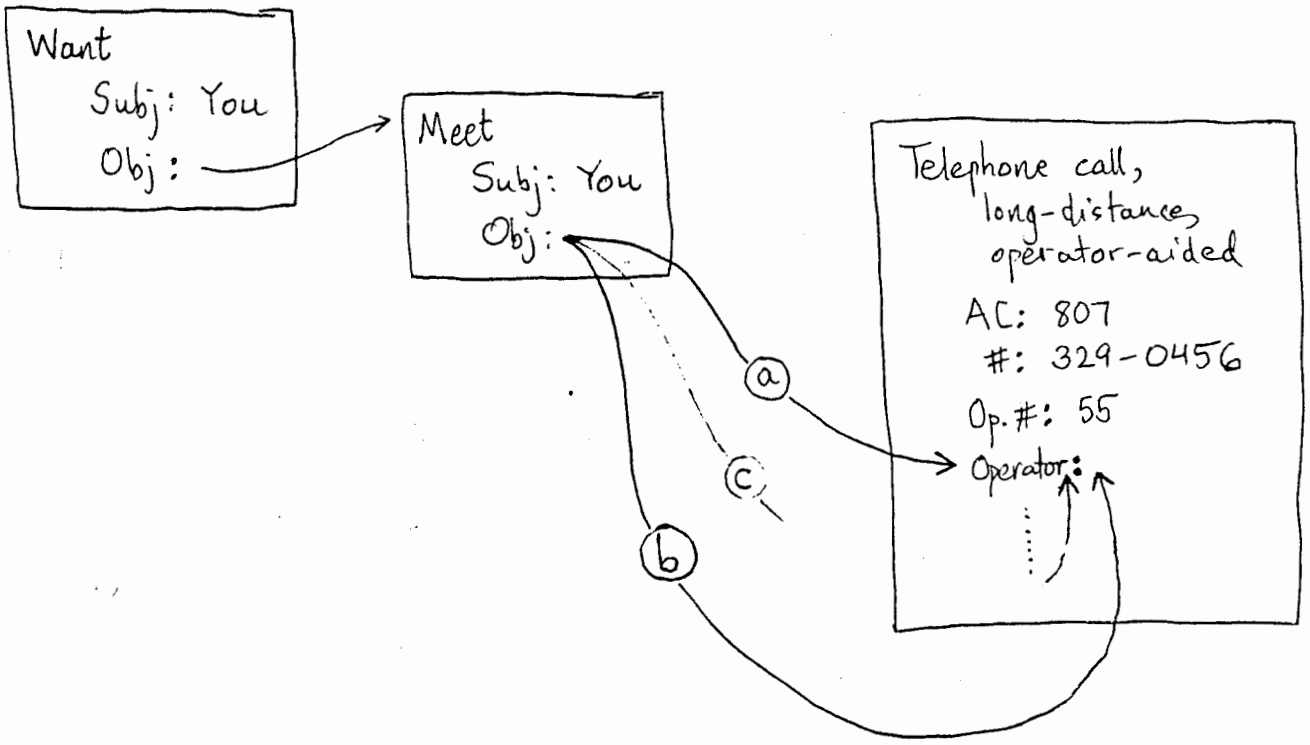


Figure 4. Three ways of thinking about meeting Operator #55.

ROLE your desired object fills. But suppose she quit and someone else filled that role -- certainly your desire should not remain rigidly attached to the role. Therefore, Figure 4a does not correctly represent your desire.

Nor does Figure 4b, with its pointer going to the right of the colon. This faulty version implies that you know exactly who it is that you want to meet, that you have some sort of direct access to her, it's just that you've CHOSEN to describe her as "Operator 55". Actually, you have no choice, because that's the ONLY way you know her -- yet it's not her intrinsic, defining characteristic. Figure 4c shows your plight accurately. You want to meet someone who you can describe only as "Operator 55", but you want to meet her for another reason entirely. These types of subtle distinctions are extremely important, and will recur throughout this paper.

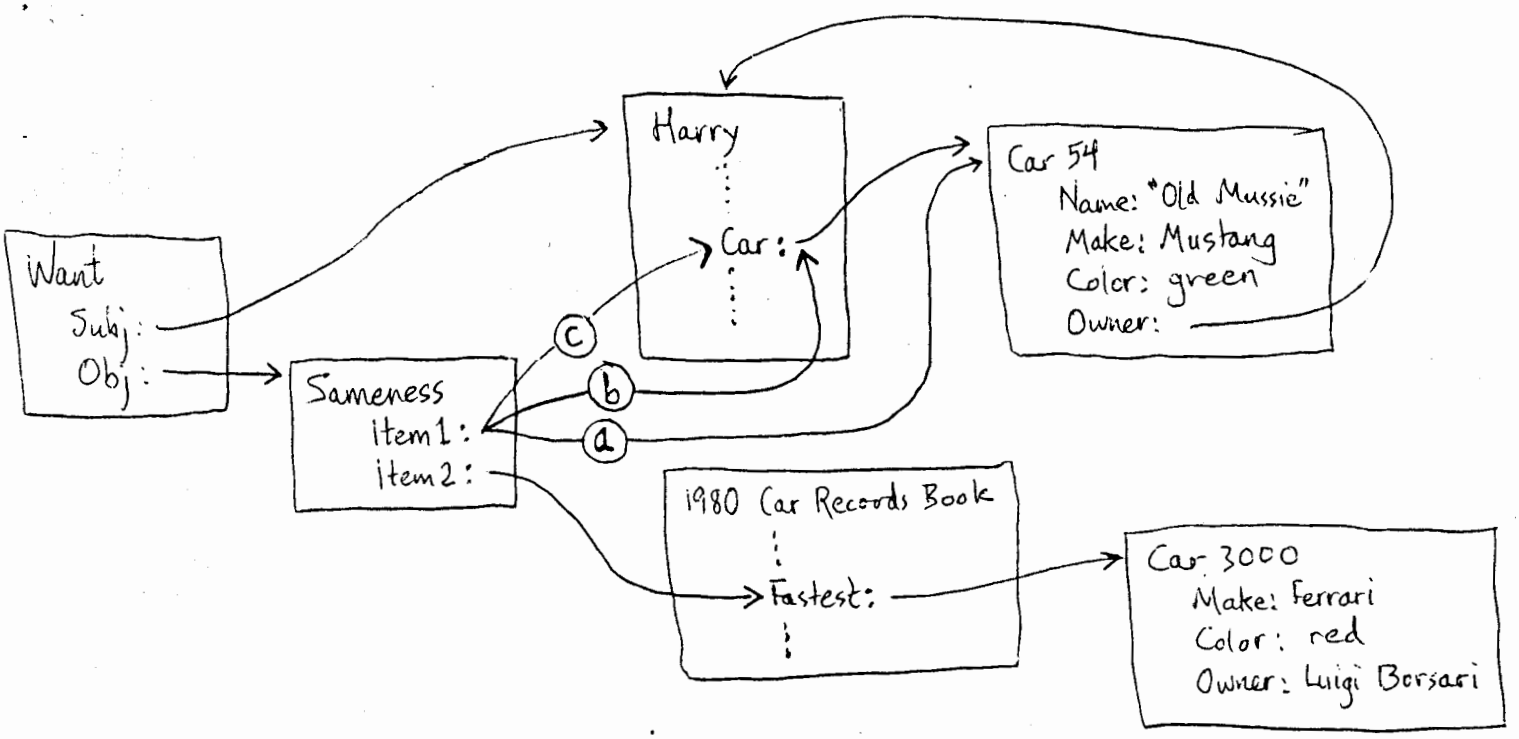


Figure 5. Three different versions of Harry's wish for his car to be the fastest car in the world.

Let's go back to the fastest car now. In Figure 5, we have shown what happens with three different versions of the pointer to Harry's car. Version (a) is when the pointer goes directly to "Old Mussie". This one would be verbalized by Harry as, "I want Old Mussie to be the world's fastest car." In version (b), the pointer goes to the right of the colon in the "Car" slot of the "Harry" frame. This one would be verbalized as, "I want my car to be the world's fastest car." But so would version (c), where the arrow points to the slot itself! What, then, is the difference between these two? Well, in version (b), Harry is THINKING of Old Mussie but he's too embarrassed to call her by name, so he just DESCRIBES her as "my car", whereas in version (c), he's not thinking of Old Mussie at all -- in fact he wants WHATEVER CAR HE OWNS to be the fastest in the world. Maybe he'll go down to the Porsche store and trade in Old Mussie (horrors!) for a new Porsche and soup it up so it'll beat that red Ferrari! Version (c) here is referring not to a physical car, but to a ROLE once again -- here, the role of "car owned by Harry".

DUMMY NODES

In Figure 6, we have taken version (b) of the previous Figure and shown another way of representing it. This one involves creating a new node, a "Dummy Car" node whose raison d'etre, or reason for creation, is simply to fill the role of "Fastest". Thus its "Core ID" slot points back to that slot, to justify its existence. (Strictly speaking, the Core ID pointer is pointing to the QUOTED slot, for otherwise it would mean that the Core ID is a CAR when in fact it's a SLOT.) We have already seen some dummy nodes -- in Figure 3, Girl-1, Man-2, and Woman-2 were dummies; their Core ID's were omitted for simplicity, however.

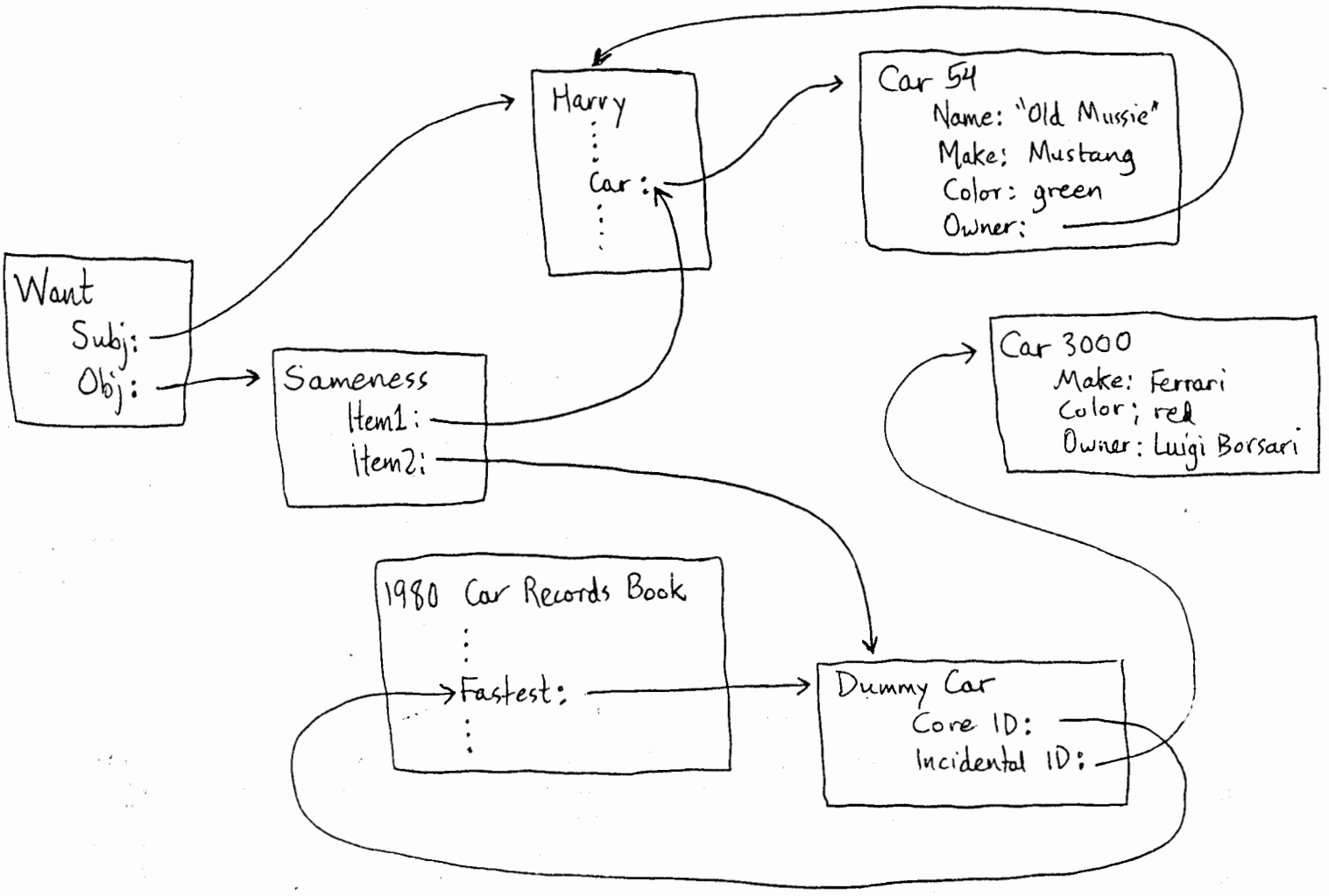


Figure 6. Harry's desire, represented using a dummy node.

In our "Dummy Car" node there is another slot, the "Incidental ID" slot, whose value happens to be "Car 3000", which is the node for the red Ferrari. Harry may or may not know the fastest car in the world is a red Ferrari. But it is obviously ridiculous to think that Harry wants to turn green Old Mussie into a red Ferrari! Harry wants his car to be the slot-filler for "Fastest", and that is what Figure 6 expresses, with or without the pointer to the red Ferrari -- or rather, the NODE for the red Ferrari! (It's so easy to slip and confuse nodes with external objects!)

So far, pointing at a dummy node would seem to have no advantage over pointing at the slot it's defined to fill. So why create it at all? To see, let's go on to Pete, in Figure 7. Actually, Figure 7 is wrong, but let's see why. If you examine it, you'll see that its proper English rendering is, "I want to be the owner of Car 3000, that red Ferrari." But who says Pete has ever heard of this red Ferrari, or of Luigi Borsari? All he wants is to fill the "Owner" role for whatever car happens to fill the "Fastest" role.

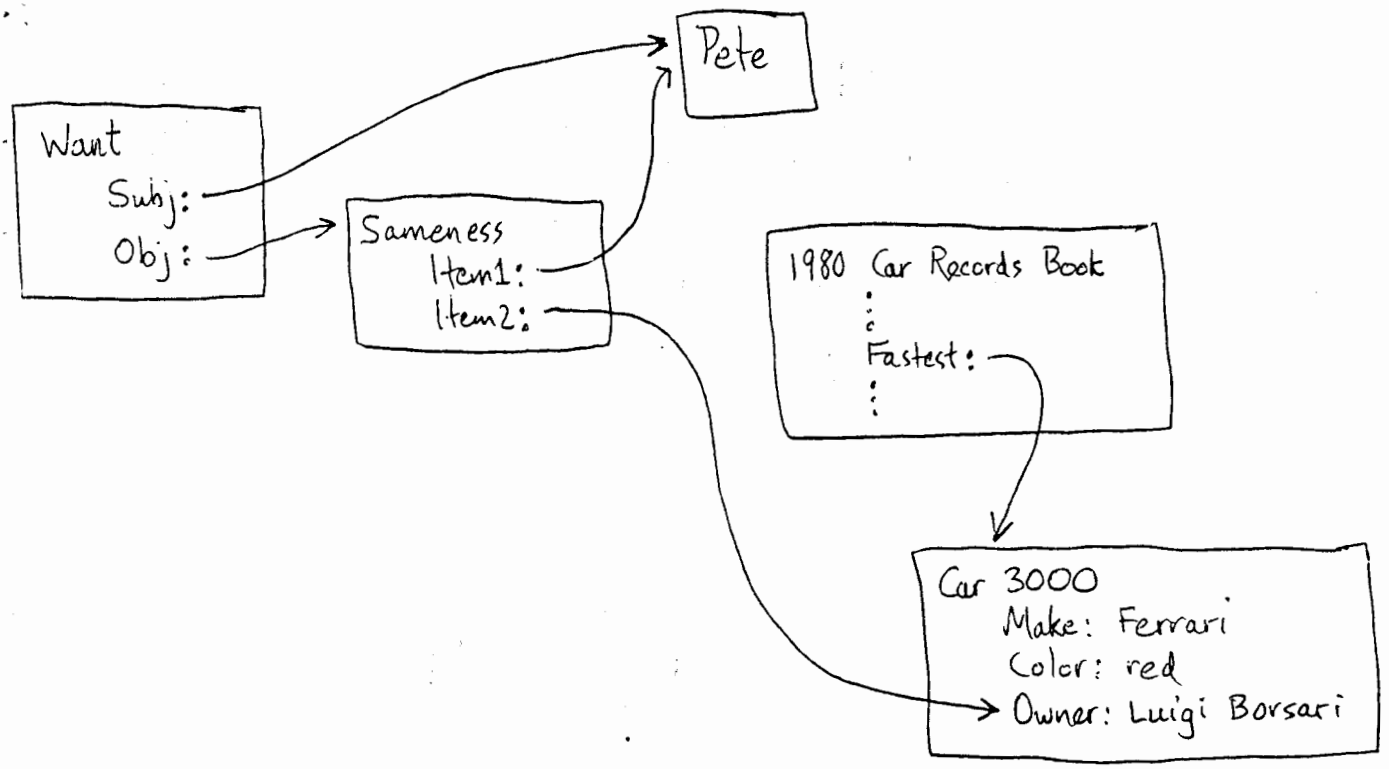


Figure 7. A misrepresentation of "I want to be the owner of the fastest car in the world."

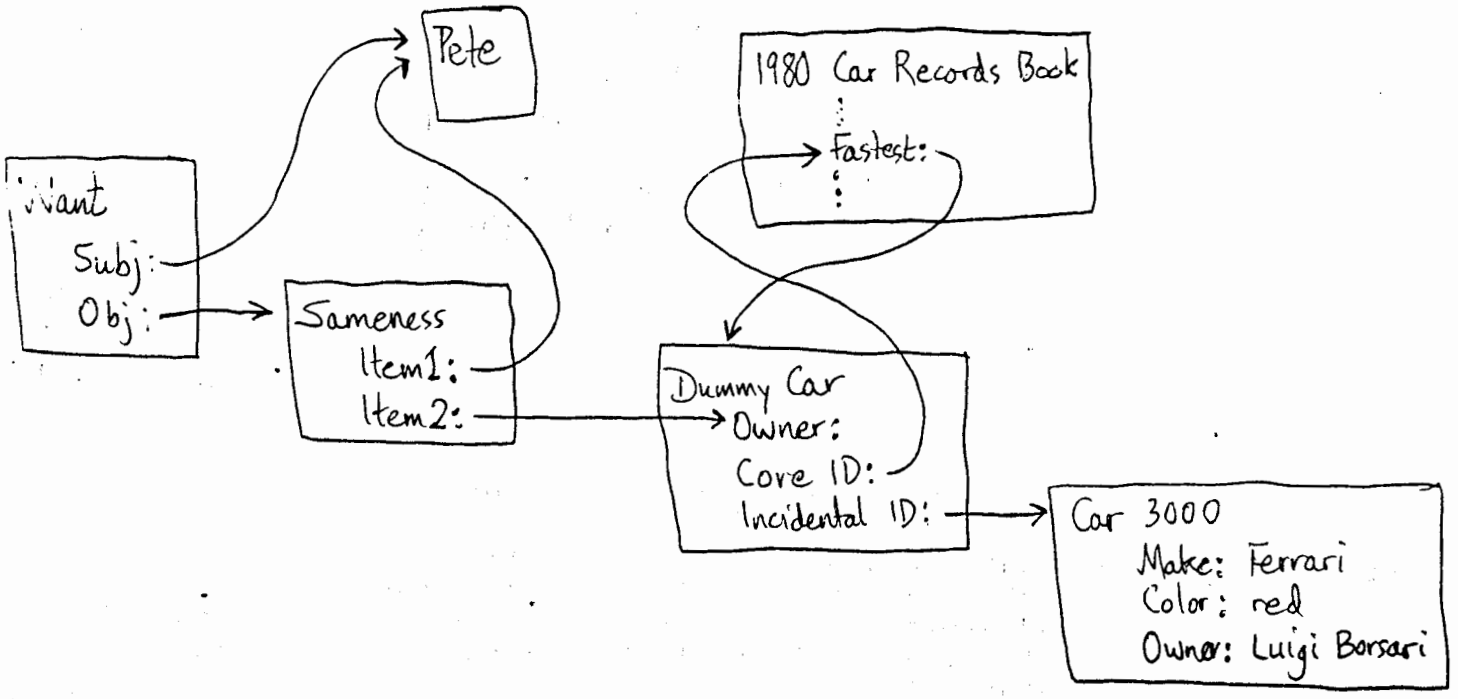


Figure 8. "I want to be the owner of the fastest car in the world" -- correctly represented, using a dummy node.

Now our dummy node comes in handy, because we can give it an "Owner" slot. It is a little like having an "Owner of Fastest Car" slot in the record book. Giving a slot its own slots can be achieved only through the use of a dummy node. An improved diagram is shown in Figure 8. The reader might think about how this diagram differs from one that would express what Bartholomew means when he says, "I want to own the fastest car in the world." What HE means, underneath it all, is "I want to be Luigi Borsari!"

In Figure 9, we have given a ghostly sort of embodiment to the operator with the seductive voice. Now we have a "Dummy Operator" node whose Core ID is that it fills the "Operator" slot in the phone

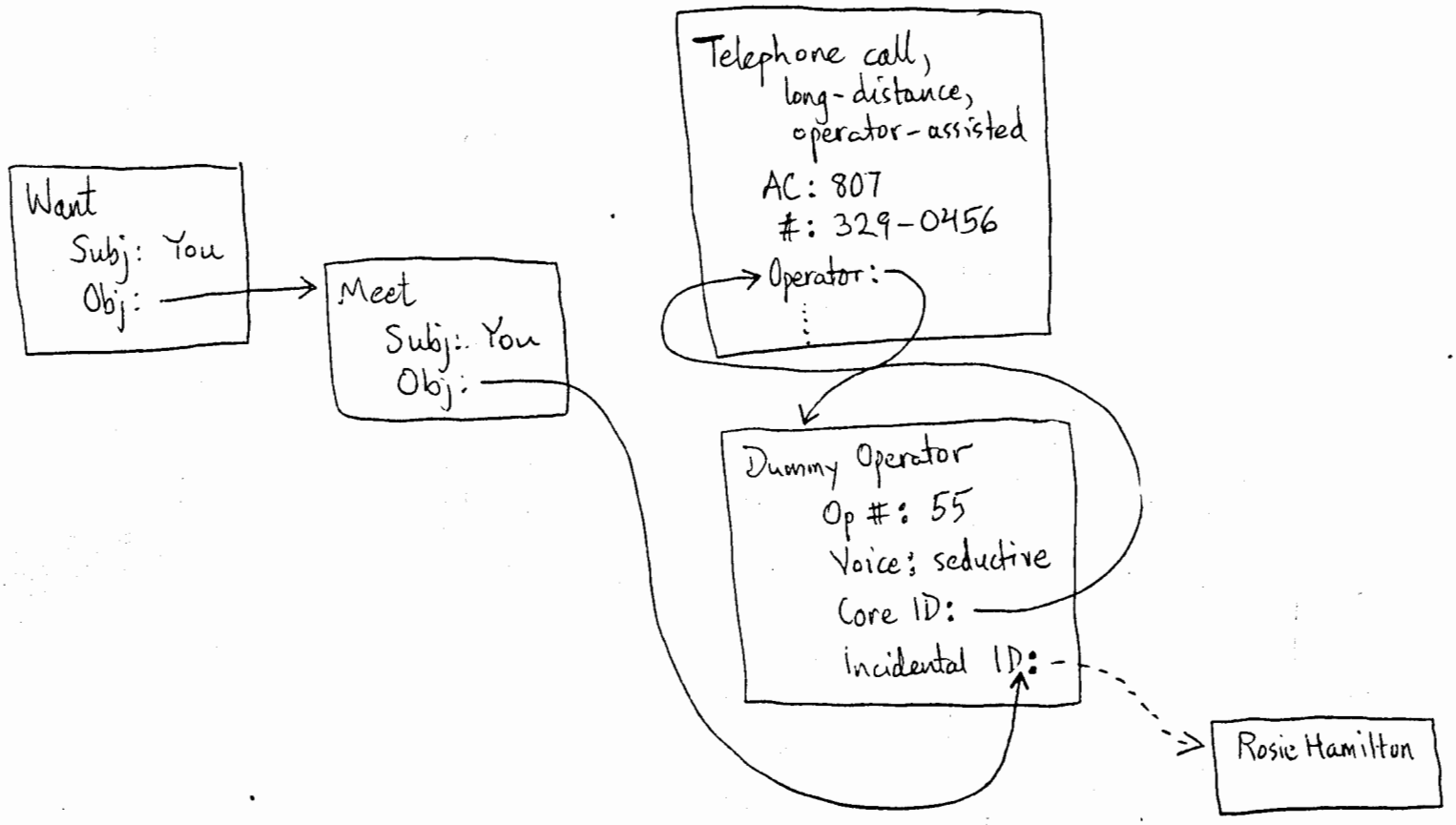


Figure 9. Operator #55 represented as a dummy node.

call. We might say that you don't want to meet "the operator", but the "person who she is". The language is a little awkward, but it makes the point -- and explains why the pointer is now directed towards the "Incidental ID" slot (to the left of its colon, of course) instead of towards the "Operator" slot, as before.

WHEN IS INTENSION PREFERABLE TO EXTENSION?

We now consider how we would represent the role "fastest car in the world" in the absence of any record book. The idea is, of course, that in a giant race of all the cars in the world, this car would win. (We say "of course", but in fact, it is not always the case that to find the "fastest" of some category, you make a race -- for instance,

consider "the fastest gun in the West", "the fastest of the nine planets", "the fastest shutter speed", "the fastest color film", etc. And if someone spoke of "the fastest apple in the orchard" it would be quite unclear how to think about it, just as it would be if someone spoke of "the saddest car in the world". A process taking into account both the words "fast" and "car" is required here -- no mean feat in itself, but not the central concern of this article.) However, obviously, no such race will ever take place. Perhaps a better way of thinking of the meaning of the phrase is that, in any hypothetical TWO-CAR race between this car and any other car, this car would win. Figure 10 shows a diagram which gets at this notion.

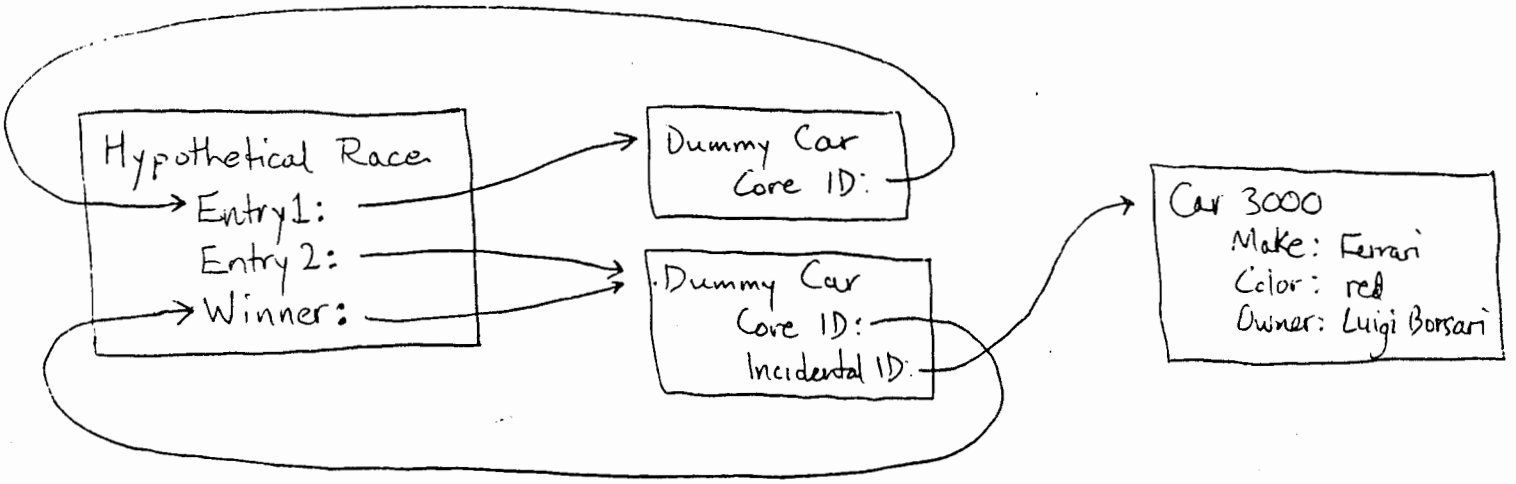


Figure 10. A representation of "fastest car" halfway towards being procedural.

Consider, however, the differences between this rather abstract, global situation and a more concrete, local situation, based on the phrase "the tallest block in the box". Whereas the car-race situation doesn't call for action but rather implies a kind of daydreaming, one may actually want to pick up the tallest block in the box. Thus, one can't merely imagine hypothetical two-block tallness competitions! One needs a diagram that describes an ACTION to be taken or a PROGRAM to be run, to find out the identity of the block filling this role.

It is not so hard to convert the idea of hypothetical two-way competitions into a program that will carry them out. What may be a little harder for an artificial intelligence system is to know which situations call for actual, runnable programs to determine the slot fillers, versus situations that are only to be imagined. Take sentences like these: "It wasn't the funniest novel I ever read, but it was pretty funny anyway", "She must have been the reddest-haired woman in Europe", "We went swimming in the river on the hottest day in the history of San Antonio", "Even the stupidest kid in my class could solve that problem", and so on. None of them calls for actual pinpointing of the extensions (i.e., the slot-fillers), although most of them could be identified if need be. Their value is that they conjure up vivid images. The question raised here is, "When do you look for the extension? When do you remain content with the intension?"

GEORGE MEETS THE HEAD OF THE DEPARTMENT

We now switch to another story, this one about two hypothetical characters named George Bilhoolie and Ben Galtiger (Figure 11). One day graduate student George said to Doctor Ben, his thesis advisor, "I'm going to have a meeting with the head of the Philosophy Department."

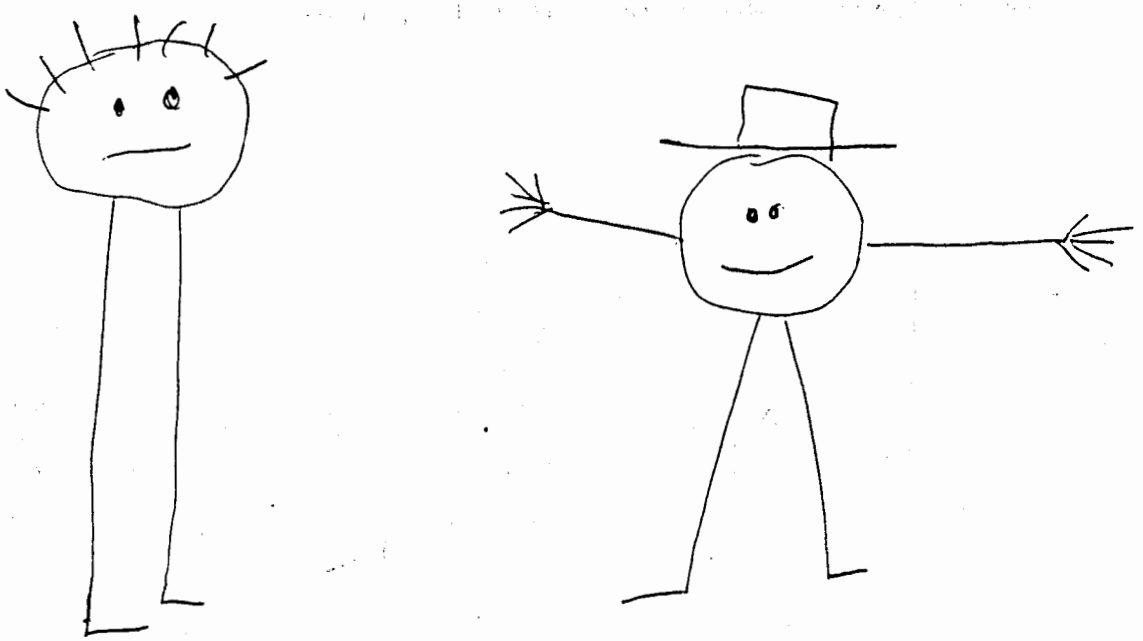


Figure 11. Ben Galtiger (right) tells George Bilhoolie (left) that Ebenezer Goodge (not pictured) has never heard of him (George).

A few days later, Ben asked, "How did that meeting go?" George replied, "Oh, very well." Ben said, "Hmm, that's funny. I talked to Ebenezer Goodge, and he said he'd never heard of you!"

Figure 12 explains the whole misunderstanding. Look at what George had in his head when he first was talking to Ben. In English, it is the idea., "I'm going to talk with the head of my minor department." But this is too awkward and formal-sounding for a casual remark. Besides, George and Ben had recently been discussing what George's minor would be, and the latest thought along those lines was philosophy, so why not just say "philosophy"? Ben would know what he meant.

Well, a couple of days later, George revised his opinion about his minor for the umpteenth time, and decided to make it psychology instead. So when he actually called up to make the appointment, he called the PSYCHOLOGY department office, and set up an appointment with Bella Gulosi, head of that department. And that's the appointment that he thought Ben was asking about. Now how in the world could George think THAT, when he had told Ben he was going to have an appointment with the PHILOSOPHY department head? Easy -- George had no recollection at all of the way he had described the forthcoming appointment, because choice of words is a pretty minor thing. All he remembered is that he had told Ben about the appointment he was going to have. In George's mind, there was ONLY ONE PERSON involved all the time -- the head of his minor department. Bella and Ebenezer were "the same person", so to speak.

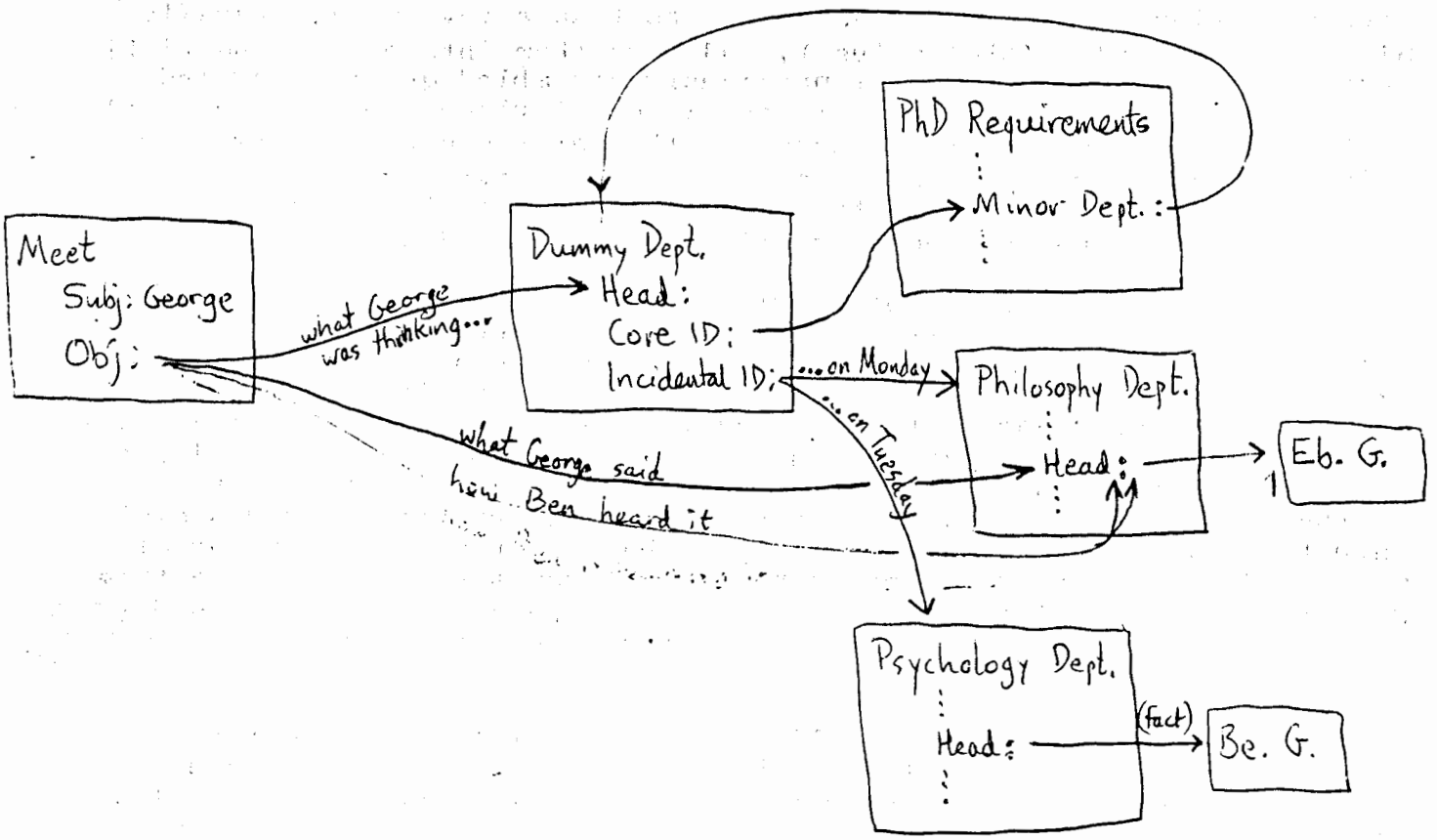


Figure 12. The misunderstanding between George and Ben, resolved by means of frames and pointers.

For his part, Ben didn't remember any better how George had described it. Perhaps George had said, "Philosophy department head", perhaps he'd said "Ebenezer Goodge". But all Ben committed to memory was this image of George talking with wizened old Ebenezer. And thus the misunderstanding, and its resolution by means of pointers and frames!

SIMULTANEOUS MULTIPLE POINTERS IN MEMORY

The final result in George's head, after the meeting, is a set of various representations for the meeting. It is simultaneously portrayed as (1) "the meeting with the head of my minor department"; (2) "the meeting with the head of the Psychology Department"; (3) "the meeting with Bella Gulosi"; (4) a vivid visual image of her actual office, the building, the location of her office, its windows, bookshelves, and so forth, Bella's face and demeanor. The first three are clearly amenable to a representation involving frames and slots; it is doubtful, however, that the last is anywhere near feasible within any present-day representation scheme.

It is true that Minsky's original formulation of the "frame" idea [Minsky 75] was partly motivated by a desire to get at visual imagery, and that part of the idea was that a visual image is what you get when you fill in some slots of a prefabricated frame, the remainder of

the slots having their default values. The idea here is that great representational power comes when you can take a few small, specific bits of information (slot values), and plug them into a vast and highly organized network of chunks of previously assembled general knowledge (frames with many defaults). All the latent implications of the simple slot values become directly accessible through the giant frames with which they are now connected. And these ideas have given rise to inheritance chains and so forth. Yet we feel that the promise of visual imagery through frames is still a long way off.

But let us return to the simultaneous multiple pointers. They raise two extremely interesting and critical issues. (1) HOW do we live with several variant descriptions of an event? Can't inconsistencies crop up? (2) WHY do we retain several variant descriptions of an event? George's four versions of the meeting with Bella range from being highly abstract, indirect, and intensional, to being highly concrete, direct, and extensional. Why do we often substitute one description for another, either out loud, as George did, or in our heads, as Ben did? Intuitively, we are perfectly aware of the different implications of various distinct descriptions of people or events -- so why do we then do violence to our intuitive judgments, and characterize something or someone in an inappropriate but possibly shorter or simpler way?

An example of (1) follows. Pete was about to take a shower, but he was expecting a phone call from Frank. Pete hates to hear the phone ring when he is in the shower and can't answer it, so he has the odd habit of bringing his phone into the bathroom and placing it on the counter so he can answer it from the shower. However, he generally prefers not getting any phone calls while in the shower. So this day, he had just moved his phone into the bathroom when he remembered Frank was going to call, and it would be annoying to be in the shower expecting a phone call any minute. So he decided to call Frank and tell him not to call back for the next fifteen minutes. By habit, he walked into the bedroom to find the phone. When he saw a blank space on the table where it usually is, he thought to himself, "You dummy! The phone's in the bathroom! You just put it there for your shower! That's the whole reason you're calling Frank now!" Pete had simultaneously remembered and forgotten that his phone was next to the shower, because he had two pointers in his head to the location of the phone. Yet he survives. Interaction with the real world corrects his course when he gets too far off.

Or how about this... You want to find a book (or turn off an air-conditioner, or something) in a very dark room at night. You start out to turn on a lamp to see by, but it's so dark you're stumbling over one thing after another. So you back off, and go for a closer lamp that will allow you to see your way over to the original lamp. You turn it on, feeling quite self-congratulatory over your clever stacking of goals ("Hey, I'm not so dumb, am I?"), and now in good light, head back to lamp #1. You're about to flick on the switch when this sheepish feeling sweeps over you... "Wait a minute... What am I turning THIS light on for? The room's ALREADY lit up!" So, a little deflated, you go and do your original task. Hasn't something like this ever happened to you?

FLEXIBILITY vs. EFFICIENCY: THE TRADE-OFF BETWEEN RAISED AND LOWERED POINTERS

In these real-life vignettes, people convert a description of what they want to do into a MECHANICAL PLAN, with all roles replaced by "lowered pointers" -- pointers that, instead of pointing at ROLES (such as purposes or place descriptions), point directly at FILLERS (such as desired effects or place names). By contrast, a "raised pointer" is one that points to a ROLE rather than to the NODE that fills the role.

The process we call POINTER LOWERING amounts to a conversion from slots to fillers, or roles to nodes. In this process, things get more concrete, more direct, or "reified". In a sense, one gains information because all things are completely identified, but in another sense, one loses information because their functions are forgotten. The converse process, which we call POINTER ELEVATION, amounts to abstracting out roles from nodes. In this process, things get more abstract, more indirect -- perhaps less efficient but certainly more flexible.

In any case, we often have multiple internal representations of events, actions, people, etc., whose conflicts we do not sense until some critical moment when they come to the surface. THERE IS NOTHING WRONG WITH THIS! This kind of "stupidity" is an important feature of intelligence! It is a sign of tremendous flexibility.

Much work in artificial intelligence is concerned with efficient sorting of subgoals, making sure there are no conflicts, redundancies, and so forth. This is called "planning". We feel this is an overly mathematical approach to intelligence. We are much more impressed with Sussman's "Hacker" [Sussman 1975]. This big "metaprogram" could write little programs to accomplish various tasks in the blocks world. Its beauty and depth came from the fact that not EFFICIENCY but FLEXIBILITY was the concern. Many people misinterpret the purpose of Hacker, thinking that its major purpose was to have a computer itself create efficient blocks-world programs. How much more impressive it is to make a program that can (even if gropingly and haltingly) debug its own faulty programs, than one that simply writes perfect ones right off the bat!

WHEN DOES SOMEONE'S NAME ACTUALLY MEAN THEIR ROLE?

Here is an example of issue number (2). Ben's department secretary says to him, "You'll have to get this signed by Paul." She means Paul Jones, his department head. Why does she substitute the individual -- the role FILLER -- for the role itself? Why not say "by the department head"? Why the lowered pointer?

One could mention a number of reasons. First, it's shorter. Second, probably the department head won't change before Ben gets it signed, so her using his name rather than his title is unlikely to steer Ben wrong. Third, it would sound funny to refer to someone by his title when they both know him -- as if she thought Ben didn't know know Paul by name. Fourth, common sense (or rather, common knowledge) tells Ben that signatures are seldom required of specific INDIVIDUALS, but very often of ROLE-fillers, so he himself can figure out that she means "Paul, in his capacity as department chairman." If, on the other hand, she were to say, "You'll have to have this form signed by Mitch" (another professor), he would wonder why, and his mind would immediately begin searching for some special ROLE that Mitch fulfills! (This kind of knowledge about signatures and roles could be stored in the "signing" frame, and would become accessible when the verb "sign" is processed.)

A similar but subtly different example of a confusion between role and filler occurred when Ben called George on the telephone. George's roommate Henry answered, and Ben took his voice for George's. They quickly sorted it out, and George got on the line. Ben said to George, "Hi, George. I thought Henry was you." A picture of this is presented in Figure 13. Ben used the name "Henry" as a shorthand in order to convey the idea of the ROLE "Telephone call answerer".

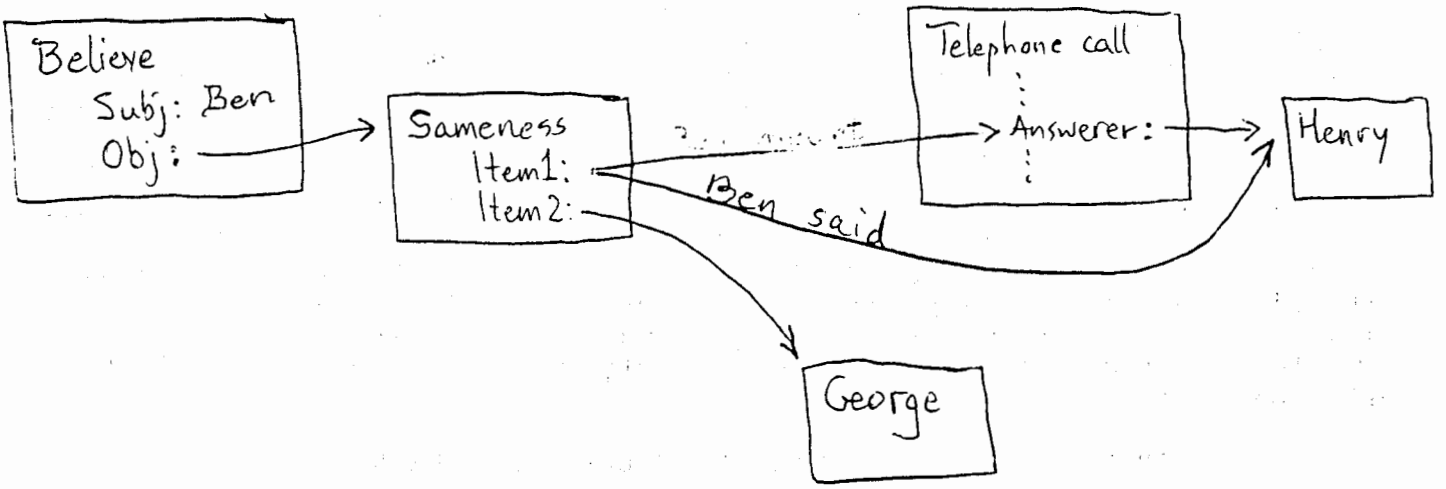


Figure 13. Ben is less confused than he sounds.

It would be strange indeed if Ben had fused in his mind the nodes for the two PEOPLE George and Henry! But it would also be strange and awkward for Ben to say, "Hi, George. I thought the person who answered the phone was you", when he knows perfectly well that the name of the answerer was "Henry".

What these examples seem to suggest is that people almost always prefer to refer to someone by name rather than by role, even when the person's role is what they mean. It seems as if people prefer to use lowered pointers -- at least in the public act of speech, if not in the private act of thinking. Somehow, it seems preferable to communicate via the concrete image of a person rather than via the abstract and impersonal image that a description by role creates. This must be somehow related to the idea that a name is the closest we can get to a true extensional representation of a person.

It is hard to come up with a case where things would be clearer if a speaker used a DESCRIPTION of someone rather than their NAME when the hearer and speaker both know the person well (and each knows that both know that person). It's not so hard to come up with cases where social protocol, idiomatic usage, emotional pressure, or some other external constraint would justify using a description, but how would it ever be CLEARER to refer to someone you know, not by name, but by role?

The only case we've come up with is a hypothetical one like this. Imagine a smallish high school where everyone is required to take both music and P.E., and where there is only one music teacher and only one gym teacher. One day these two teachers get together and the gym teacher asks the music teacher, "How's that symphony coming?" The music teacher replies, "My first flutist isn't quite up to her part, but otherwise it's coming pretty well." He didn't say "Alice Blaine" despite the fact that he knows the gym teacher knows her, because he was thinking of Alice, THE FLUTIST, and besides, the gym teacher still might not know Alice plays flute. "How's the girls' basketball team shaping up?" he goes on. The gym teacher replies, "Well, my center is great, but the rest of them aren't hustling enough." She, in turn, didn't say "Alice Blaine" because she was thinking of Alice, THE CENTER -- and didn't know if the music teacher knew Alice was center (after all, Alice isn't very tall!).

Actually, even this example isn't quite what we want, because although both people know Alice, they don't both know they're talking about her. Is there a case where both know "who" they're talking about, and yet still use a description by role?

FACETS

This idea of visualizing some specific individual, yet specifically AS A ROLE FILLER, is an intriguing one. If we think back to Ben's statement to George, "I thought Henry was you", we can imagine that Ben's between-the-lines meaning was: "I thought Henry, WEARING HIS HAT AS TELEPHONE ANSWERER, was you." This suggests a division of Ben's mental image of Henry into "facets". Ben assumes that George likewise will retrieve only the obvious facet of his representation of Henry, the facet of Henry as telephone answerer.

Without the notion of facets, our representation would have to be as was shown in Figure 13. George retrieves an indivisible Henry-node from memory, then traces back from it to a recent slot it filled, and thus construes Ben's sentence. This is similar to the way that was suggested for interpreting the secretary's sentence, "This form has to be signed by Paul." The crucial features here are that a node is INDIVISIBLE (i.e., is retrieved wholly or not at all), and that the roles it plays are EXTERNAL to it (they are slots in other frames).

With the notion of facets, the picture changes. The division between a role and the node that fills it is blurred, for a node can be retrieved PARTIALLY (i.e., by facet), and each facet represents a single role (or collection of related roles) that the node fills. There may still be slots in other frames which point to that node, but tracing pointers backwards to find those slots is no longer necessary. This picture would explain why people tend to use names even when they are referring to roles: they hope to induce, in the hearer's mind, the retrieval of a node representing some individual, but only a PARTIAL retrieval, i.e., a retrieval of some facet or other.

Politicians often take advantage of the reverse effect. The chairman of the Republican party gets on television and makes a big statement about some issue, and then adds, "Of course, I am not speaking as the chairman of the Republican party, but merely as a private individual." Swell... but if this is so, why should THIS private individual have gotten TV time in preference to millions of others? Strangely, few people ask this question, because for most people it is hard to retrieve just the private-citizen facet of a well-known person -- they can't help summoning up the public-figure facet along with it.

It would be beyond the scope of this article to suggest ways of implementing either of the two alternative schemes -- that is, with facets or with no facets (or some hybrid scheme). The main point is that this is a significant distinction which should be considered carefully.

CONCEPTUAL SKELETONS -- THEIR DYNAMIC PRODUCTION AND ELUSIVE CHARACTER

The process of retrieving a node by a role it plays, and the converse process -- the creation of a fresh role to fit a node -- are exceedingly mysterious aspects of human thought. Perhaps the latter, though, is the real mystery. A couple of examples will show why.

Take this commonplace sort of interchange:

> Sam: I haven't written my parents in a long time.
> Sue: Me neither.

Clearly Sue is referring to HER parents, not to Sam's parents. Yet the reference Sam made was exclusively to his own parents. How can Sue get away with this sloppy statement "Me neither"? It happens all the time! What is going on? We'll come back to this later, but first let's see another fragment from a typical conversation.

> Henry: It took me a while, but I finally found some stamps in my wallet.
> Kathy: My husband's wallet is a real mess, too. Oh, well, I should talk -- you should see my purse!

What does Kathy's purse have to do with Henry's wallet? Why did she bring it up? Because to her, Henry's comment about his wallet created an abstraction of sorts -- the generic notion of "messy wallet", which is not attached to any specific owner. Thus her husband's wallet, falling in this category, was retrievable. Once she had mentioned it, it became further abstracted, into some sort of category -- we call it a CONCEPTUAL SKELETON (see [Hofstadter 1979]) -- that is vaguely verbalizable as "messy small-items-carrier". This conversion of a node for a SPECIFIC object into a description of a CATEGORY of objects is, to us, a central mystery of human thought. It is closely related to pointer elevation, described above.

One reader of the above paragraph pointed out that Kathy's conceptual skeleton probably included the idea of "something which you almost always have with you". We concur, and feel it only strengthens our belief that these conceptual skeletons -- categories that are continually produced "on the fly" by human brains in the most ordinary of situations -- are very elusive when it comes to trying to capture them in words.

To show that this phenomenon is closely related to roles and role-fillers, how would you explain the following? Doug went to graduate school for several years in the small university town of Eugene, Oregon. A few years later, he got a job in the small university town of Bloomington, Indiana. From the start, Doug would catch himself on the verge of saying "Eugene" when he was meaning to refer to Bloomington. This was annoying because he certainly felt he did not confuse the two towns in his mind. What was weirder, though, was one day when he caught himself about to say "Portland" when he meant "Indianapolis". This happened more than once. And Doug also noticed the reverse phenomenon occasionally, where he would catch himself about to say "Bloomington" when meaning to refer to Eugene.

What is going on here? Somehow, the nodes for the two small towns are stored close to each other, probably in a common "mental compartment". They play very similar ROLES in Doug's mental representation of the world. But the Indianapolis-Portland confusion is probably more revealing. Here we have two cities, neither of which Doug knows well, both of which, in his mind, are "satellites" to the small towns he knows. Each of them is, roughly, "the big, mostly boring city an hour or two's drive to the north". It may or may not include such features as "don't know well", "site of the University's medical school", "internationally not very significant". Doug also lived in Stanford for many years and yet he never confuses San Francisco with Portland or Indianapolis; he knows it somewhat, and does not find it boring (incidentally, it used to be site of Stanford's medical school!). San Francisco does not fill the analogous role in Doug's Stanford-frame that the other two cities do in their respective frames.

The interesting thing to notice is that there is certainly no generic, prepackaged "small university town" frame containing a "big, mostly boring city (etc.)" slot ready to be filled in every time Doug moves to a new university town. The characterizations of Portland and, later, Indianapolis were manufactured dynamically and independently of each other. They are certainly not English phrases, but rather, some sort of pointers to the nodes for the cities. It is obviously false that Doug's characterizations of Portland and Indianapolis are IDENTICAL, yet somehow the addressing mechanisms in Doug's brain still manage, on occasion, to confuse those two pointers, and thus to retrieve the "name" slot of the wrong node.

There is something beyond the descriptions themselves, something more abstract -- something like the nonverbal "messy small-items-carrier (etc.)" conceptual skeleton in Kathy's brain -- that is responsible for this curious wire-crossing. We feel it would be a hopelessly complex task to try to model this in full at this stage. But we can attempt a simple and limited approximation to it. To begin with, we define a notion we call ROLE, which is a generalization of the formal concept of "slot".

ROLES, AS FRAMES, HELP EXPLAIN ANALOGICAL THOUGHT

Certain frames -- ones we call "ROLE-FRAMES" -- may function something like slots. When such a frame appears as a slot name, it does not function merely as an atomic identifier. Instead, it indicates that information inside the role-frame is to be used in creating a pointer, or sequence of pointers, that eventually locate the slot's filler. So, finding the filler of a slot in a frame can involve more than just following a pointer labeled by a certain atomic identifier. It may require looking up the "meaning" of the slot name in some structure -- in fact, in what we call a role-frame.

For example, how would you explain this kind of an exchange?

American: Carter is sure messing things up here.
Canadian: Yeah, we're not too happy with Trudeau either.

It's simple enough -- one is tempted to call it trivial. Yet it tacitly involves the solving of this analogy problem:

?????????????? is to Canada
as
Jimmy Carter is to the United States

In attempting to solve this problem, one might begin by tracing various pointers backwards from the "Jimmy Carter" node. Because of its saliency, the "President" slot in the "United States" frame would presumably be found quickly (along with the fact that Carter fills it). However, there is no slot by the same name in the "Canada" frame. As if this weren't bad enough, though, consider this next one:

American: Carter is sure messing things up here.
West German: Yeah, we're not too happy with Schmidt either.

West Germany has, it happens, both a President and a Chancellor. The "obviously correct" solution to the analogy would be to take the West German President, right? Yet none of the authors of this paper was sure who the President is. Perhaps someone named Jenschke?

In West Germany, the President occupies a more ceremonial post, while the Chancellor is the prominent political figure. Nearly anyone interested in politics, whether in the United States, Canada, or West Germany, would name Pierre Trudeau and Helmut Schmidt in a flash as the counterparts to Jimmy Carter. Few of us would stop even for a moment to ask the question, "Let's see, does West Germany have a position of 'President'?" Many people who could do the analogy would not even know Trudeau's or Schmidt's (or, for that matter, Carter's) title.

How do we do this, if not with some mediating notion along the lines of "President"? Well, certainly there must be something like that, but it is more abstract than the literal WORD "President". And substituting the rigid phrase "Head of State" is not going to get us much further, either. (In this particular case, it might solve the problem, but it is not pointing in the direction of a general solution to the problem.) What we need is a more abstract kind of slot, which can function at more than one level of abstraction, and this is what we call a ROLE, represented in a ROLE-FRAME. If the "President", "Prime Minister", and "Chancellor" slots were represented by ROLE-FRAMES rather than atomic slot names, then, rather than comparing mere identifiers, one could look inside the role-frames and see that the three perform the "same job" in their respective "parent" frames. Just how this kind of knowledge would be stored in a role-frame is not clear; we do not claim to have this issue licked!

A similar but considerably harder analogy problem would be this one:

??????????? is to Great Britain
as
Rosalyn Carter is to the United States.

We easily came up with four candidates, all arguably "right". If Rosalyn Carter is simply "The Most Venerated Lady in the Land", then her counterpart might be either Queen Elizabeth or Margaret Thatcher. On the other hand, if Rosalyn Carter is seen as "Wife of the President", then would not "Wife of the Prime Minister" (following the Canadian example) be the analogous role? The problem is, of course, that Margaret Thatcher has no wife, but a husband. "Wife" would have to be converted into "Husband", and then we would retrieve Mr. Thatcher. There was once a newspaper article about Denis Thatcher in which he was variously referred to as "First Lady", "First Man", and even as "the quintessence of dutiful political 'wives'". The fourth possibility depends on seeing Rosalyn Carter's role as "Wife of the Foremost Man", which gets converted without too much difficulty into "Spouse of the Foremost Person", and this can then retrieve Prince Philip for us.

The "First Lady" concept is a stretchable but complex one. What goes on in your mind when you try to make sense of these phrases: "First Lady of General Motors", "First Lady of the ACM", "First Lady of the Vatican", "First Lady of orchestra instruments", "First Lady of feminism", "First Lady of barnyard animals" -- and so on?

It is clear that a generalization of the rigid notion of slot is needed, and this is the "role" idea. Moving upwards in abstraction from a slot to its role is analogous to our previous notion of pointer elevation". Now we can slip a pointer from a node to a slot, then to a role!

How is a role represented? Consider the "First Lady" role. In its basic form, it means "Wife of President". This is a CHAIN of two slots. This is the typical way a role is defined -- as a chain of slots. Thus, the role-frame must define that chain.

In Figure 14, we've shown a role-frame plugged into two distinct contexts. The first example is an encoding of the newspaperish sentence "The First Lady [of the United States] visited an orphanage," where the phrase "First Lady" is used simply for variety, and really means "Rosalyn Carter". The second example is an encoding of the sentence "The First Lady [of France] lives in Paris," intended as an eternal verity.

In the Figure, you see the role-frame with its chain of slots. Also there are wavy arrows pointing off to the general concepts lurking behind the slot names. The structure of that abstract knowledge is symbolized fuzzily by a cloud of related ideas. This is not meant to be taken literally! It is simply to show that such knowledge is accessible from the role-frame. The organization of such knowledge is a thorny problem -- one to which we offer no solution at this time. However, we should mention here that Ronald Brachman and others, coming at some of these questions from a different angle, have been developing many interesting and pertinent ideas on roles and intensionality in their work on the language KL-ONE [Brachman 1977, 1978, 1979a, 1979b].

SCENARIOS, CONCEPTUAL SKELETONS, AND ADDRESSING

Now let us go back to consider Sue's statement, "Me neither." She meant, "I haven't written MY parents in a long time either." What did she abstract out of Sam's statement? Clearly a notion something like "person writing to own parents" -- perhaps more complex, in that it involves adverbs and so on, but this is enough for us. This phrase is not like a role in the sense of "chain of abstract slots", for it involves a combination of a person, an action, and slots in the nodes that represent them. This is more like a small scenario, so we call a structure representing a generic type of event built up out of abstract actors and actions a "scenario". (It may seem that such a scenario is something like a "script" in the sense of Schank, but a scenario is considerably less intricate than a script, and is dynamically manufactured.) A diagram of the scenario which Sue created in her mind on hearing Sam's statement is shown in Figure 15.

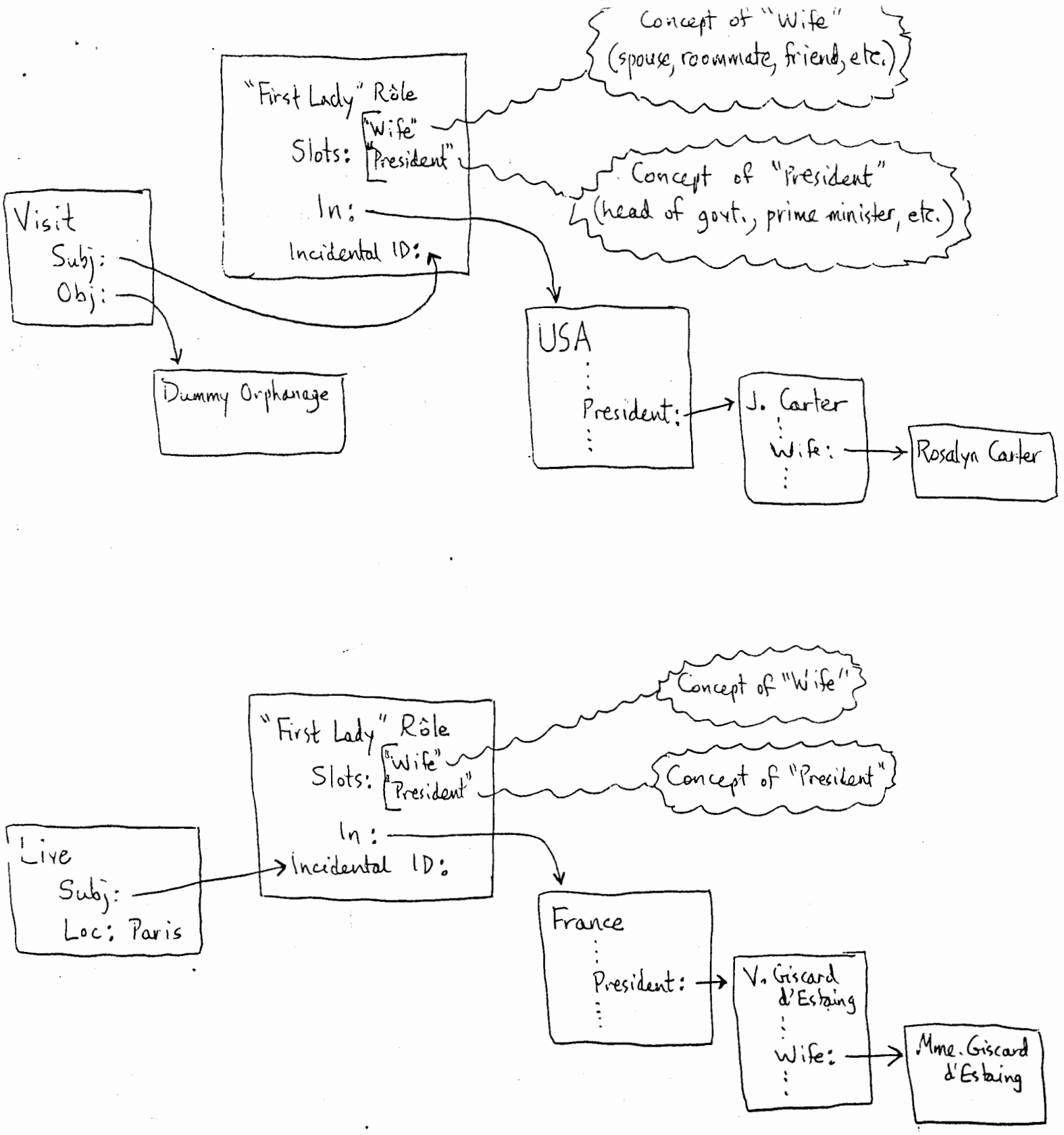


Figure 14. How a frame for the "First Lady" role can be plugged modularly into two contexts. The upper diagram represents the sentence "The First Lady visited an orphanage," while the lower one represents the sentence "The First Lady lives in Paris."

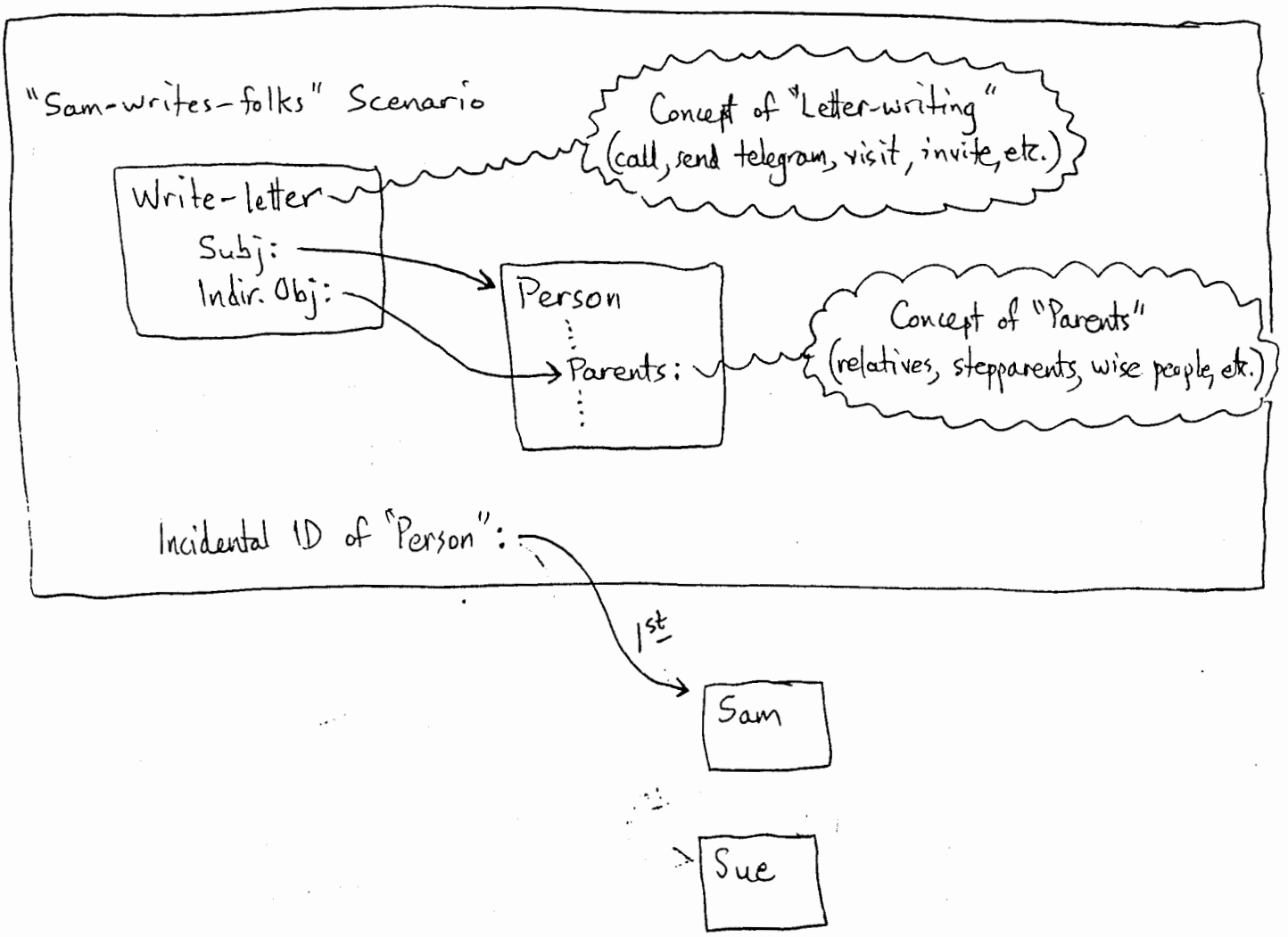


Figure 15. Sue's scenario (simplified) upon hearing Sam's sentence.

It must be borne in mind that this scenario, no less than the "First Lady" role, is more flexible than its surface appearance. We could just as easily understand Sue's train of thought if the dialogue had gone this way:

Sam: I haven't written my parents in a long time.
Sue: Oh, no! I was supposed to call my godmother last night!

Here, her underlying conceptual skeleton is clearly something like "feel sense of guilt towards older loved one(s) for having too little contact". Yet it is surely not so patly verbalizable. There is something elusive about it which we feel will challenge knowledge representation workers for a long time to come. The notion of "scenario" represents our first attempt to get at the conceptual skeleton which a person creates when hearing a full sentence.

An example of something rather akin to a conceptual skeleton that occurs in the AI literature is the "abstract (teleological) process model" of Sussman, in his book about the program "Hacker" [Sussman 1975]. In the Chapter on how Hacker debugs the little programs it writes, Sussman shows how Hacker extracts, from a completely annotated trace of a run with a bug in it, an abstract model which contains only the essence of the situation (at least it is the essence along one conceptual dimension!). This skeletal version of the process then serves as a "key" into a knowledge base of programming errors, from which heuristics for rewriting the program can be drawn. The elegant thing about this is that Sussman's conceptual skeleton serves as a sort of "abstract addressing mechanism", which is precisely what we believe conceptual skeletons are.

A standard use of conceptual skeletons as "addresses" is when someone tells a joke, and it reminds you of another joke that you know -- that is, the first joke is converted by your unconscious mind into an abstract skeleton, and that skeleton is used as a key which dips down into your memory and retrieves a second joke which shares the skeleton. The similarity that links such jokes together is often quite abstract, having little to do with their overt subject matter; it is a challenge then to try to capture the elusive essence in words. These kinds of ideas on conceptual skeletons are connected with some of the ideas that Roger Schank brings up in his work on the phenomenon of "reminding" [Schank 1979], and with Dedre Gentner's notion [Gentner 1980] of structural maps in her studies of analogies in science and literature.

INTENSIONALITY AND PLACES

Much of what we have said about intensional descriptions of PEOPLE pertains as well to PLACES, TIMES, and other entities with unique identities. One has to wonder, for instance, when one visits Greenfield Park in Dearborn, Michigan and is told "this is the place where Thomas Edison discovered the light bulb" -- especially if one knows that he discovered the light bulb in Menlo Park, New Jersey. The people who made the park somehow got the HOUSE in which the discovery was made transported from New Jersey to Michigan. Apparently some people feel comfortable with the idea that the house IS the place; to them, there can still be a little thrill at standing by the table on which he did his work, and so forth. What do they feel when they look out the windows, though? Then again, would you or I feel any more historically accurate if we went and stood in the parking garage (or whatever) which now stands on the old site of Edison's house in Menlo Park, and thought, "THIS is where the light bulb was discovered"?

Examples like this come to mind quite easily, once you put yourself into the proper frame of mind. A restaurant moves to the other side of the street; a chef moves from one restaurant to another; anything that characterizes one place gets attached to another place... The house you were born in and grew up in still exists but is now

surrounded by ugly and decaying suburban sprawl instead of a pleasant small town feeling. All of these can tear up our nodes and cause, if not mental anguish, at least real disorientation and some anxiety. The whole thing has to do with how we nest frames of reference, and how, in various contexts, we (unconsciously) give different amounts of weight to various frames of reference in determining where we are. For example, "Where are you?" will be answered in extremely different ways depending on whether you are talking to an astronaut, your little brother in his room, someone on the other end of a phone, etc., etc. The reader may find other examples of place-identity quandaries in [Dennett 1980].

We often refer to places by analogy: for example, a New Yorker might describe Hyde Park as "London's Central Park", or Piccadilly Circus as "London's Times Square" (or "London's Columbus Circle"). Here, the exact function of the place in the unmentioned city is unspoken, but presumed understood by the hearer -- i.e., turned into a role of some sort, abstracted into a "portable" feature of a city. But neither speaker nor hearer makes explicit what that role is -- and perhaps would be hard put to do so!

A strange place-identity phenomenon happens in subways or trains. If you want to be at the NORTHERN end of the platform when you get off at 116th Street, and you're standing waiting for the train at the SOUTHERN end of the platform at 59th Street, you have various options. You can wait till you get out at 116th Street and then walk from one end of that station to the other; you can walk right now inside the 59th Street station to the north end of the platform; or you can even do your walking once you're inside the train. In some sense, you will take "the same walk", consisting of "the same steps", no matter which choice you make.

THE SOULS OF ELECTRONIC WORDS AND LETTERS

Curiously, the experience of text-editing on a computer affords some rather bizarre experiences involving identity and intensionality. It all has to do with what you consider to be "the same character". Nearly everyone who has used a good screen-oriented text-editor will recognize the following sorts of sensations. Suppose I type a line and then tell my text-editor to copy that line below itself, so I have two identical lines. Now I delete the original line, and the copy jumps upwards, replacing the original. Every character is the same, but are they THE SAME characters? As programmers and scientists, we may scoff at such questions, but the children and animists in us will nonetheless feel a kind of emotional tug when we delete some favorite remark and replace it with a mere copy of itself.

There are a host of related sensations. Who hasn't scrolled a group of lines right off the top of the screen, then scrolled back down and felt, "There! The very lines that were floating up above the screen came back down and there they are again!" Certainly one has the feeling that a sentence which is being "pushed" rightwards by characters being inserted to its left is THE SAME SENTENCE, not a copy! What if you type the same sentence right on top of itself on the screen? Is the result a copy, or is it the original?

A strange effect is produced when you type out, at high speed on a screen, a long file all of whose lines are identical. The screen seems to stand still, as if nothing were happening! But of course, WE know that REALLY, all the lines are moving upwards... Or are they? If the screen is standing still, then nothing is moving -- or is it? It is rather confusing, because the dot matrix is being refreshed at the same rate whether the file is standing still or being scrolled. The "motion" is behind the scenes, in the operating system and the communication channel to the terminal. To complicate matters slightly, have each line be ALMOST the same as the one above it. Then MOST of the screen stands still -- but in a narrow vertical strip, surrounded by an immobile background, a stream of characters snakes its way upwards and off the top... One can get very confused about what "same" and "different" mean when watching a screen!

THE SOULS OF PEOPLE

All of this can perhaps be placed in a more serious-sounding context if, instead of thinking of characters on a screen (which, after all, we know are mere shadows of a deeper "reality", which consists of strange magnetic patterns rotating at breakneck speed on a disk in some far-off air-conditioned building), we think of PEOPLE. Let us indulge ourselves, and imagine that copies of people can be made in a science-fiction-like device called a "human-editor". If YOU were copied and there existed two atom-by-atom identical copies of you, and one of them were about to be exterminated by someone pushing the "DELETE" key on a "human-editor", would you care which one it was to be? Or would you (plural) cavalierly dismiss the worry as "academic", "childish", "irrelevant", "animistic" -- and flip a coin? After all, as long as the Grand Cosmic Disk in the Sky stays spinning with all souls on it, who cares which superficial manifestation gets deleted from the Screen of Life? To some people at least, this is a very serious matter coming straight at the question of the existence and location of the human soul. In that sense, the decision about characters on a screen is a metaphor for one of the most profound questions in all human existence! Rather than being a ridiculous frivolity it now seems to resonate with cosmic import!

SHAKESPEARE'S PLAYS WEREN'T WRITTEN BY HIM...

With this grandiose prologue, we hope now to have created a properly weighty atmosphere in which to broach the title sentence of this paper: "Shakespeare's plays weren't written by him, but by someone else of the same name." This strikes us as obviously humorous, but what is really going on in our minds that makes us smile?

We shall first consider a rather simple-minded theory. But to do so, we need some general notions that will be helpful also in describing more complex theories. So to begin with, let us assume that we all have a node in our minds (or in our AI representational systems) that stands for a set of works including the plays "Hamlet",

"Macbeth", and so on. Let's call this node "Works-1". Under Works-1 we have all sorts of entries, such as play names, sonnet names, measures of greatness, fragmentary excerpts such as "The quality of mercy is not strained," and so on. We have also a node for a certain human -- actually an author, in fact a Great Author -- so let's call that node "Great-Author-1". This node has some slots such as "Name", which is of course filled with "Shakespeare". Under Great-Author-1 we have also references to April 23, 1564, Stratford-on-Avon, Ann Hathaway, etc. Of course the two nodes are also mutually intertwined, each receiving definitional quality from the other. That is, each fills a slot in the other. (See Figure 16.)

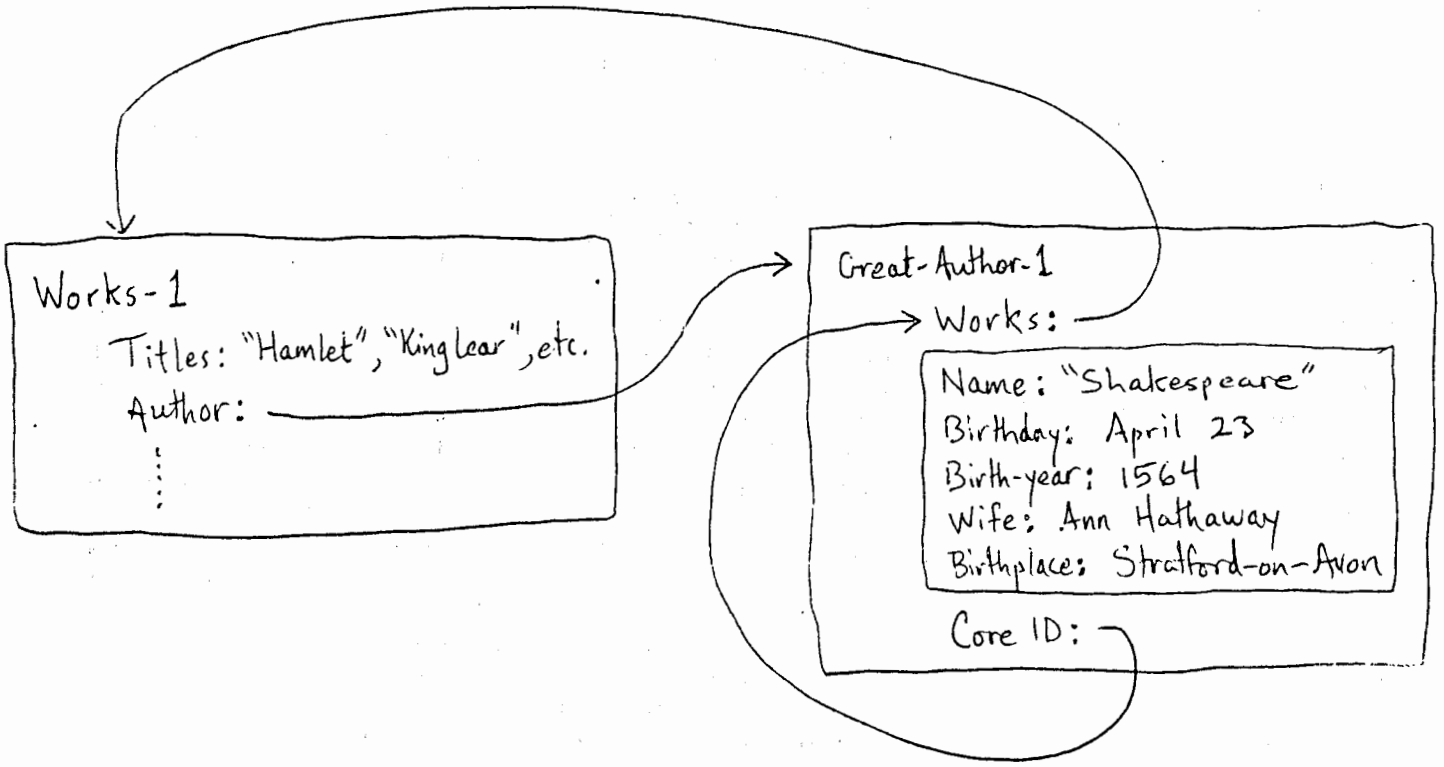


Figure 16. A naive representation of the connection between Shakespeare and his works.

Now here's the simple-minded theory. It suggests that when we hear the first clause, we do nothing other than reach into the frame Great-Author-1 and wipe out the word "Shakespeare" as the value of its "Name" slot; then when we hear the second clause, we merely reinstate it as the value after all! Thus the sentence as a whole is simply a fancy no-op -- no wonder we laugh when we hear it! There is a degree of reality to this view, but we have to go beyond it. Could it be that when we first hear this sentence (knowing that it is a joke), this is the first thing that happens?

And now let us consider a more embellished theory. It seems certain that, as we start to process this sentence, we are induced to retrieve the node Great-Author-1 by angling in at it through one of its slots, namely the "Name" slot, filled by "Shakespeare". (It is presumed here that a name has a back-pointer to its frame.) From Great-Author-1 we follow the "Works" pointer over to Works-1. Thus, in some sense this retrieval has "activated" both nodes -- that is, made them focuses of attention.

We now read, "weren't written by him". Who is this "him"? What role are we defining by using this name? We have a dilemma. Up until reading this clause, we had had a single person whose identity involved a number of components: being the author of certain plays, having a certain name, having a certain birthplace and life story, and so on. Now we are in essence being told to split up this single personage into two pieces: one retaining the name, birthplace, date, etc., and the other retaining the authorship of the famous works Wl. Which of those two pieces will we mean henceforth, when we use the name "Shakespeare"? We are somewhat torn between two ID's.

Well, common sense tells us that a name, as well as other down-to-earth aspects such as birthdate and so on, belongs to a corporeal, flesh-and-blood person, while a role-filler definition such as "author of such-&-such works" creates in our minds a less human image, one more ethereal and abstract, one to which a name would be much more incidental. Thus, it seems that we unconsciously choose to have the name be retained by the "impostor" fragment -- that is, the one with only the birthday and other "trivialities" -- whereas (until the second clause) the other fragment has no attached name. So we are more or less "forced" to retrieve the impostor, when we hear "by him" (which we take as synonymous with "by Shakespeare").

When we split the old node up into two pieces, we must either have new nodes coalesce around the two fragments, or let one fragment survive and prosper as a full-fledged, healthy frame, while the other simply hangs in limbo, discarded and unwanted. This seems a little too cruel! So we -- along with most people -- will create a new node so that both characters may be retained. Let us give a name to our new node. Well, which IS the new node? Should the new node represent the new author, or the impostor?

It seems that we should try to disturb the structures attached to the old node as little as possible, simply so as to minimize the amount of pointer-shuffling we'll have to do. It seems pretty easy to make the "Author" pointer of Works-1 point off to a new node, so let's say that Great-Author-2 is the node that Works-1 now points to. In other words, Great-Author-2 is a new "Great Author" frame about which we know little, besides the fact that its "Works" slot must be filled with a pointer to Works-1. We don't yet know the name and other "trivial" features of Great-Author-2. Still, it has room for them, having sprouted blank slots for them the way a sliced earthworm will sprout new appendages. These slots are inherited from the class "Great-Author".

As we link up Works-1 and Great-Author-2, we simultaneously deactivate the "Works" pointer from Great-Author-1 to Works-1 and the "Author" pointer going in the opposite direction. All this leaves Great-Author-1 rather deflated, now representing a mundane personage, born in Stratford-on-Avon on April 23, 1564. (See Figure 17.) Note that the old pointers are deactivated but not totally destroyed.

Psychologically, the dilemma (at this point) is that hearing the name "Shakespeare" causes us to retrieve a historical nonentity, someone whom we'd never have heard of were it not for some quirk of fate. Yet we have known of this person for years, and attribute to him

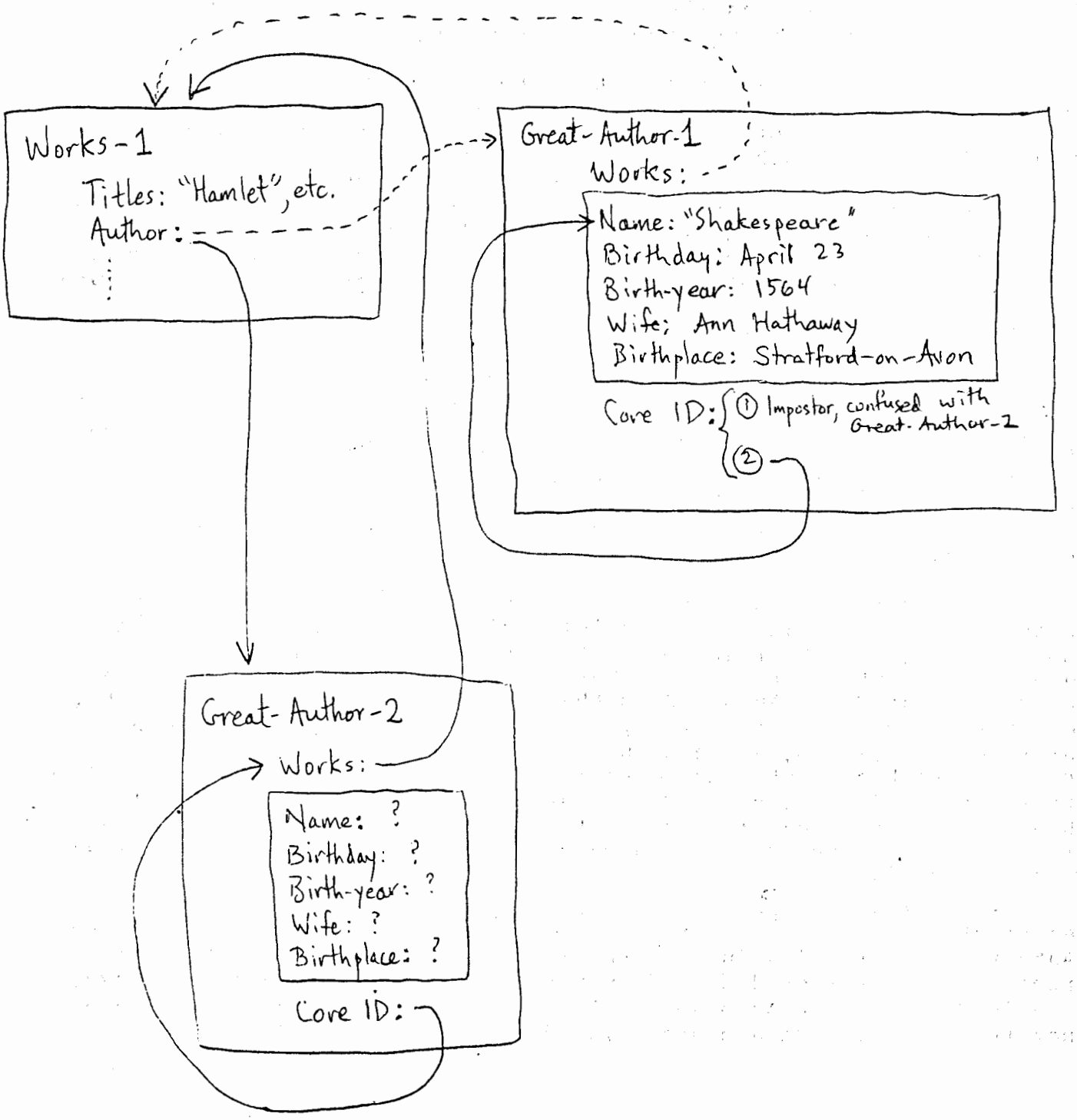


Figure 17. A representation of the state of knowledge of a person who has heard the first clause of the title sentence.

some sort of "soul". We don't want to relinquish this fact; we don't want to impersonally liquidate a soul simply because he didn't write some plays -- yet the strange thing is that that soul is a "friend" of ours only because we mistakenly assumed he was the author of those plays. It is as if we met someone by accident, some sort of mistaken identity in a hotel lobby, became fast friends, then discovered that our basic presumption about this person's identity has to be totally undermined. Or even more painful, it is as if someone we know has undergone a tremendous personality change, such as can occur after a brain-damaging auto accident, or as a result of senility. How do we deal with such a soul-wrenching situation? "Shakespeare" for us still conjures up a certain mental portrait, along with a painful "impostor" label -- not that it was HIS fault. Perhaps simply "wrong fellow" is better than "impostor".

Until just now, the Core ID of the person known as "Shakespeare" for us resided in his authorship of certain plays. Now that has been taken away from this person, and someone else has this as HIS Core ID. (Here we mean something a little different from the earlier meaning of "Core ID", when it pertained only to dummies, unlike Shakespeare. What we mean here is something like "key identifying characteristic".) What is left for poor Shakespeare as his Core ID? Unfortunately, it is simply this "mixed identities" fact -- that, plus the fact that he has the name "Shakespeare", an important name in our mental repertoire. Note that strictly speaking, we can no longer even assume that this Shakespeare, still represented by the node Great-Author-1, is a Great Author. He may be just some schmoe, who never wrote a thing. Yet we still want to retain the (deactivated) pointers connnecting it with Works-1, since they are an important part of the history of how we came to think of this schmoe. Untimely ripping out the "Works" slot would be too drastic. Somehow we must have the flexibility to undo the "Great-Author-ness" of the node Great-Author-1, yet to retain vestiges of its original pointers.

Though not yet at clause two, we still have one further conceptual difficulty to clear up. We have a seeming contradiction. We have used the phrase "Shakespeare's plays" to get us via a certain route to a certain node (Works-1), and then in the same breath, we have undermined the validity of the route. This is not actually a contradiction, though it comes perilously close to being one. In essence, we are being told, in a sort of shorthand way, that the phrase "Shakespeare's plays" can no longer be used to retrieve a certain node -- but amusingly, the very act of retrieving the node to attach this fact to it is done, for the final time, by the about-to-be-forbidden route. At this point, we have roughly realized a structure which encodes this idea: "The plays which up till now we have thought of as being by someone named Shakespeare are no longer to be considered so. They are by someone else about whom we know very little."

... BUT BY SOMEONE ELSE OF THE SAME NAME

Now let us go on to the second clause. Clearly, when we hear it, we fill the "Name" slot of Great-Author-2 with a pointer to "Shakespeare". Now we can attempt to describe the humor. As in the simplistic theory, we seem to have done a lot and yet achieved practically nothing. Why has very little been achieved? Largely because we knew very little, to start out with, of the original Shakespeare -- so the new Shakespeare is not all that different from the old. We have practically restored the original state, except that we have cluttered up our model with a silly extra node. The situation has been restored to the extent that the supposedly invalidated intensional pathway to node Works-1 -- the phrase "Shakespeare's plays" -- has been revalidated.

The disorienting feature of our new state is this. "Shakespeare" as an author's name now designates a different node, a different person. It still directs us to the same work, Works-1, so it seems to play the same role (in fact it DOES play the same role!) -- it just is attached to a different "soul"! That is, we know that our old William Shakespeare, to whom we'd attributed a certain soul, now has been preempted. We're using another node to function where he did, and since soul and role-cluster are usually so closely related, it confuses us. We're being asked to separate out, distinguish between, two things we don't usually: (1) the soul, or personal identity, and (2) the set of roles fulfilled. For a stimulating discussion of ideas of this kind, see [Kripke 1980]. Another forthcoming book on related matters is [Hofstadter-Dennett 1981].

Let us consider, for the sake of further clarification, the case where our sentence has not yet been heard, but I tell you, "Latest scholarship shows Shakespeare's birthday to be September 15, not April 23." You'd unhesitatingly accept that switch (unless you had a special reason to care -- for instance, if your birthday were April 23 and you were proud of it!). Your concept of Shakespeare as a person would be safe, unthreatened. This is only an incidental feature, a triviality. Yet the curious thing is that clause one of our sentence forces us to cling to precisely those trivial aspects of Shakespeare to maintain some meaning for "him". Where formerly his Core ID resided in his authorship of certain plays, now his Core ID rests in trivial facts, while someone else is "who he was".

Romantic breakups can give rise to emotional pangs arising from such soul-role confusions. One feels attached to a unique individual, yet one is forced to search for another unique individual to be "the same person" in a very deep sense. Sometimes the old lover is perceived primarily as the filler of that vital role, sometimes as an eternal individual. These are not easy to disentangle.

SHAKESPEARE AND GALILEO WERE BORN IN THE SAME YEAR

Let us consider one last variant on slipping Shakespeare. It so happens that both he and Galileo were born in the same year -- 1564. How do we represent this fact? Do both have "Birth-year" slots filled with pointers to the same atom? Yes; but this is only an IMPLICIT recognition of the sameness. What is the difference between a system that theoretically COULD notice that their birth-years are the same, and one that already HAS? How do you make the sameness EXPLICIT? Well, it would make sense to set up a "Sameness" node with two pointers pointing to the two "Birth-year" slots filled by the year 1564. (We hope it is clear why it would make no sense at all to have these two pointers pointing at the year 1564 (its node, we mean).)

But -- care is required here. Do we point to the Birth-year slots, or to the right or left of their colons? (See Figure 18.)

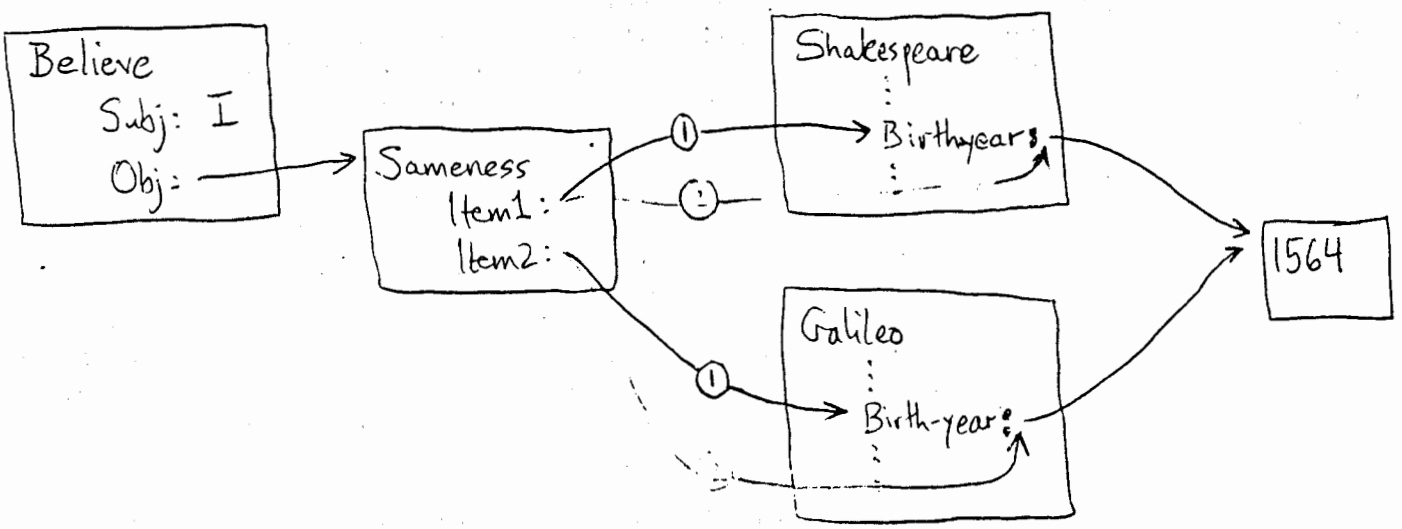


Figure 18. Two ways to represent the sentence "Shakespeare and Galileo were born in the same year."

Here's one way to figure it out. Suppose I found out Shakespeare's birth-year was 1565, not 1564. Now a pointer that points at a SLOT doesn't give a hoot if the FILLER of its slot changes -- so if they were of this type, both sameness pointers would sit merrily there, and the Sameness frame would now assert (albeit indirectly) "1564 = 1565"! Such rigidity is certainly not what we desire in our encoding. After all, Shakespeare and Galileo were born in the same year not by DEFINITION, but by COINCIDENCE! So how do we get at that notion?

What if both point to the RIGHT of the colon? The flavor of this is, roughly: "The year that, among other things, is Shakespeare's birth-year coincides with the year that, among other things, is Galileo's birth-year." (For a similar example, think back to "Jim's wife is Sally's mother.") Now suppose Shakespeare's birth-year slot value gets changed to 1565. (Jim divorces Maude and marries Nancy.) Because it is a right-of-the-colon pointer, it knows that, although it is pointing at part of a SLOT (Shakespeare's birth-year), it is really describing a NODE (the year 1564) -- so if the node gets

detached from the slot altogether, to continue pointing at that slot might be to give out false information. (Nancy is not Sally's mother, despite being Jim's wife.) So, the pointer has no choice but to inform its parent node (the Sameness frame). This frame, realizing it has been undermined, politely retracts itself. Thus, this is more like the desired encoding, because it seems to incorporate a tacit recognition of the coincidental, non-necessary character of the equality.

When two things are tied together in some way, it is always important to know if it is by coincidence or by definition (or logical necessity, or whatever). Our solution above gets at the coincidental nature of a particular sameness, but in a way which we said was TACIT. It is based on a distinction between pointers of different types, and thus is a kind of "unconscious" recognition of the coincidence. Put another way, the system IMPLICITLY understands the Shakespeare-Galileo sameness is a coincidence, but not EXPLICITLY. We wanted to make recognition of the sameness explicit, which we have done; we could now carry things further by making recognition of the accidental character of the sameness also explicit. This might be accomplished by filling in a "Character" slot in the Sameness frame, or it might be done by having each pointer carry "dependency notes" that tell what other pointers it depends on (or vice versa), and in what ways.

One could then legitimately ask whether the dependency notes themselves have dependency notes. If so, then this leads immediately to an infinite regress -- not a pretty prospect; yet if not, then we are left with a level of implicit knowledge which CANNOT be made explicit -- also not pretty. If there is a level of knowledge which cannot be reasoned about, that fact will be reflected in the system's behavior; at times it will appear unconscious, inflexible -- and, the worst epithet of all -- mechanical. Thus we certainly want at least the POTENTIAL for any implicit knowledge to be made explicit.

This can be achieved without infinite regress -- by having the system be able to FIGURE OUT about dependencies, whenever this is called for, rather than having it rely on static notes hanging around everywhere, cluttering the place up like crazy. (For an example, think back to Mike's problem in merging "Jim's daughter" (who had a cold) with "Sally" (who had the flu). He solved this by figuring out that his pointer to Sally's flu actually had a dependency -- namely on Peggy's reliability. The direct belief that Sally had the flu was then replaced by a more indirect, more abstract belief -- a sort of pointer elevation.) A dependency that had been figured out could be turned into a note and attached to the appropriate pointer, so as to obviate the possible need to recompute it. Needless to say, such a scheme would be very, very complicated; it would open a royal can of worms.

All of this talk about dependencies of pointers, and particularly about the difference between pointers tied together by coincidence and pointers tied together by definition, is reminiscent of the movie "Le Grand Amour" by the French director Pierre Etaix. In this movie,

many things are allowed to slip in humorous ways. One memorable scene shows the protagonist, a pathetic businessman, in his office. He has just hired a gorgeous, sexy young secretary, about twenty years old, and he is feeling so sorry for himself at being forty. He daydreams he is ten years younger (his age-pointer slips), and the scene metamorphoses... Here he is, vigorous, youthful, in tennis shorts, very snappy, sitting at his desk. He confidently pushes the button and in she walks -- ten years old, saying, "You wanted me, monsieur?" Parallel slippage of pointers where it was unwanted.

But what DOES slip in our minds in such hypotheticals? What should remain fixed? Why should her age remain constant while his slips? Of course, in daydreaming, there are no rules. Anything CAN slip -- what is so droll, in this case, is that the film doesn't even allow him to control his own idyll! But slipping doesn't take place only in daydreams. It takes place all the time as we try to map structures onto each other and to find parts which are analogous; it takes place as we construct mental models of the future in an effort to determine how we should act in certain situations. It would be disastrous if we had no commonsense ideas about what should slip and what should stay the same. But this would constitute another whole paper.

DOUG FEELS SORRY FOR A DUMMY NODE

Here is an another anomaly connected with Core ID's and dummy nodes. Dolores, a friend from Purdue, told Doug about a dancer who had been performing in an auditorium there a few years back and who had fallen into a hole in the stage and been paralyzed. Doug's immediate reaction was a feeling of pity. Later he tried to analyze the nature of this pity, and got rather confused. The question is, who was he pitying? In some sense, it was a person whose only identity to him was the description "person to be pitied". If you were told, "There exists someone who has had a calamitous fate befall them," would streams of tears suddenly roll down your cheeks? Certainly not. But if the whole story were told vividly enough, you might indeed cry, even never having heard of the person -- even if the person is entirely fictitious! Think of movies, where many people routinely cry.

Well then, how can Doug's reaction be justified? He is generally inclined to feel pity for any SPECIFIC person he knows who is paralyzed -- someone whose Core ID, in his world model, is preestablished. He has a sense of "who they are". But here the person's Core ID is the fact they they are pitiable, and nothing but that. How is it different from the far more abstract statement, "People are sometimes paralyzed by falling into holes", or even, "Somebody was once paralyzed by falling into a hole"? In the one case, he would instantiate no node, while in the other, he would instantiate a node. It seems that his pity has to have a node to flow out to. Yet this node really has no personal quality to it.

Like most people, Doug is by nature predisposed to feel pity for people who have unfortunate accidents befall them -- this is a procedural fact about him as a system. It seems odd that a node with vacuous identity could elicit genuine pity; it would seem more plausible that he would have the feeling, "Oh, that is the kind of situation that WOULD make me feel sad." The point of this example, in any case, is to highlight what seem to be the very subtle differences among the statements "Someone there is who is to be pitied", "Someone got paralyzed by falling", and "A few years ago at Purdue, someone got paralyzed by falling". How much "identity", "realism", "definitional quality", does there need to be before we feel SOMEONE -- a PERSON -- is there?

RETRIEVABILITY OF THE EXTENSION IS EMOTIONALLY IMPORTANT

One possible (partial) answer involves the idea of RETRIEVABILITY. Although it seems like an incidental and unimportant aspect of the story, Doug knew the place and could guess the year of the sad event. This gave him the POTENTIAL to go to Purdue and trace down the person involved. The fact that he would be very unlikely to want to do so is beside the point; the key point is that once a potential path of retrievability is visible, the node acquires a certain psychological feeling of deep rootedness in reality. Moreover, even if Dolores hadn't told him where and when it took place, Doug knew that SHE knew, and he could therefore begin to track down the person through her. We feel that this sense of retrievability is of the essence in giving nodes a deep psychological grounding.

A few more examples will perhaps intensify this idea. We might first mention, however, that we will be here treading in philosophers' waters -- dealing with the notions of "de dicto" and "de re". We are interested in finding out how we can model these distinctions in a program. We want our understanding of these issues to be so sharp that we can actually translate them into an operational language of frames, slots, pointers, roles, Core ID's, and so on.

EGBERT'S DREAM

Let us consider the case of Cristina Ortiz. She happens to be a Brazilian pianist whose picture is on a certain Villa-Lobos record cover. When Egbert saw this picture, he was so struck by a quality of her face that he bought the record (also because he happens to like Villa-Lobos). A few days later, he woke up with a clear memory of dreaming about Cristina Ortiz. Now what does this mean? Could he truly say he was dreaming about HER? He knew practically nothing about her -- not her character, not her voice -- not even her face. One photograph doesn't make you know anyone. It is tempting to say, "Oh, obviously he was dreaming about HIS IMAGE of her!" This is all fine and well, but at what stage of knowing someone does one switch over to dreaming about THAT VERY PERSON, not just an image of them? When can one actually think about a PERSON, not just have an image of them?

If your answer is, "Never; we can deal only with our images, never directly with reality," then we will argue. Certainly you can touch a person, not your image of them; you can see a person, talk to a person, talk about a person. Why not know a person? Why not dream about a person? Furthermore, even if we assume that the abstract philosophical point, "You only know images" were true, nonetheless, common language has us say, "I dreamt about X", or "I know X". But even in common language, it feels wrong to say, "Egbert dreamt about Cristina Ortiz". He just didn't know WHO SHE IS, well enough to say that he dreamt about her. Or perhaps we mean "HOW she is". (Would you be able, at this point, to dream about Egbert? About Jimmy Carter?) Where is this borderline between having an image of someone, and knowing them?

IDENTITY AND NONOVERLAPPING NODES

We would say that, although it is blurry, the borderline is crossed when enough information has been amassed in your model of them to make this person truly distinct from all other people in enough ways that operationally, realistically, you can feel this person's uniqueness clearly. Then this person can be said to have acquired a "soul" in your model of them, or perhaps a Core ID, if the term seems preferable. Shakespeare was right on the borderline, in our previous example.

This borderline idea needs a little clarification. It hinges on the idea that, in the multidimensional space that our models of people reside in, there is a certain average distance between a person and their "nearest neighbor" -- the person most similar to them. We can discuss the same issue in terms of musical style. If you know Bach well, when you turn on the radio and hear a piece that sounds like Bach, you don't think to yourself, "Gee, this is probably by someone very much like Bach!" Rather, you have a virtual conviction that it is by Bach himself. (Doug recently had such an experience. He was absolutely sure that the lovely flute and harpsichord piece on the radio was by Bach. When they announced it, it turned out to be by C.P.E. Bach -- not Bach at all, but someone of the same name!)

This kind of certainty will happen if you have a strong sense of the "space" of musical styles, and how close the neighbors are. If you do not have a well-developed sense of musical styles, your sense of the distribution of composers in this space will be vaguer, and you may confuse ones that to experts seem very different from each other. As you get to know more and more composers within a certain period, each one's style becomes clearer and clearer in your mind. Your increasingly precise knowledge could be represented pictorially by shrinking circles centered on the various individuals, which at an elementary level are so large that they overlap, but which gradually shrink until there is no more overlap. (See Figure 19.) Doug thought his Bach-circle had long since shrunk to the point of not intersecting any other composer's -- now he's eating humble pie!

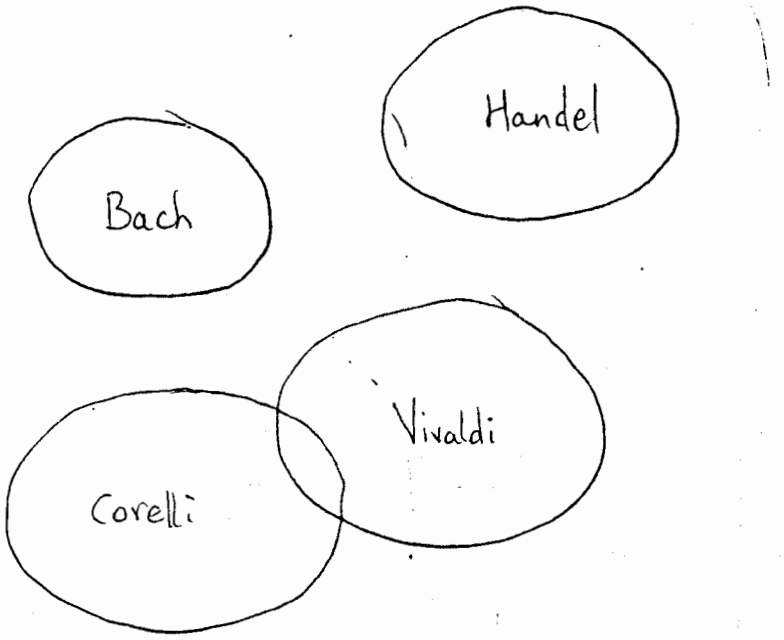


Figure 19. Overlapping green circles represent the blurring of composers' styles in the mind of a novice; smaller red circles stand for the greater discrimination in the mind of an expert.

At the early stages of getting to know any domain, one has little sense of the distribution of points in the space. This holds whether one is speaking of musical styles, artistic styles, typefaces, people, or whatever. At some stage, however, one begins to sense the density of points within the space, and then to be able to distinguish among them --·or at least some of them -- reliably. Presumably, if one really knew a space well, one could even perceive abnormally wide gaps -- unfilled niches, so to speak. Thus, one might wonder if there are unfilled niches in the space of musical instruments. Or, one might regret that no composer filled up the "obvious" niche that falls about halfway between, say, Prokofiev and Rachmaninoff.

Some of these notions might be useful to people who are interested in questions about the value of music composed very skillfully but in a long-outmoded style, the value of forgeries of Old Masters, or the likelihood that two artists of the same era would develop practically indistinguishable styles. But once again we are straying.

The point of the Cristina Ortiz example is that Egbert's model of her was too weak to count as a representation of any real person. It was too fuzzy, too indistinct -- like too wide a circle. Although she was, in a highly theoretical sense, retrievable from what he knew

about her, in any pragmatic sense, she was still irretrievable. His sense of her overlapped with his sense of too many other people. Any representational system should have strong knowledge enabling it to estimate with some confidence how well a given collection of intensional descriptions actually pins down an extension -- a real object (or person, or composer, or typeface, or artistic style) -- "out there".

THE WOMAN WITH THE RAINCOAT, AND THE PROKOFIEV BALLET MUSIC

A final pair of examples will illustrate the highly emotional responses people have when faced with intensions whose extensions seem irretrievable. Imagine that one day, David sees a woman very briefly who attracts him a great deal. As she walks around a corner and out of sight, his heart sinks. He would like to meet her, to know WHO she is. For a few moments at least, she is still accessible, in principle: if he were to run after her, chase her, and corral her. But he won't do that. And as he ponders this, she becomes gradually totally irretrievable, even in principle. Oh -- except for one fact: she was wearing an unusual raincoat which David is sure he would recognize anywhere. But he has no reason to think that he can retrieve her by that "key", because he may never see her again, or she may never wear that raincoat again... His feeling is that she exists, theoretically, somewhere "out there", but the awful thing is that, try as he will to recreate her face in his mind, he finds that it is indistinct enough that even if she turned up on his doorstep next week, canvassing for some political cause, he probably wouldn't know it was the same person. So in a real sense, she is "lost" to him. His very real longing to know THAT PERSON cannot be satisfied, the moment he gave up on the idea of pursuing her around the corner.

Sometime later, David meets an attractive woman, gets to know her, enjoys her very much, and eventually marries her. Two years later, she pulls an old raincoat out of a box, and he recognizes it. She is HER! So is that old longing, that long-dead longing, now actually satisfied? (One can invent subtle variations on this story that make it even more poignant in various ways.)

The issue may not seem all that clear yet. The final refinement of this idea is in the following real-life story. One day Doug drove down to a Chinese restaurant to pick up some hot food. As he was driving, he had the radio on and was listening to some wonderful music -- obviously ballet music by Prokofiev, but he had never heard it before. He parked in the lot and went in quickly to get the food so it wouldn't get cold. Well, it wasn't yet ready and by the time he had gotten it, paid for it, and come out to the car, of course the Prokofiev had ended, been announced, and now something completely different and of no interest to him was on. He wanted very much to know what that wonderful piece had been. Soon, its melodies faded from his mind -- he didn't write them down. If he had been desperate, he could have written the radio station or even called them up, but he didn't. Even several days later, he could have. But now it is

YEARS later, and he has no way of describing that day, even to himself, to pinpoint its date. And he would no longer recognize the melodies that that day had him so enchanted. So in any sense you want, that music is lost to him. He can never hear THAT MUSIC again.

Or can he? Why can't he just go out and buy recordings of ALL of Prokofiev's ballet music? Here is where this example is even sharper than the raincoat one. For in the case of the mystery woman, the world -- even just David's town -- is too big ever to allow an exhaustive search. It is, for all practical purposes, infinite. But in the case of Prokofiev ballet music, the universe is not just finite, it is totally accessible! (Let us assume that the radio station didn't just happen to have the world's only recording of some virtually unknown Prokofiev ballet!) So, if Doug buys all this music, plays all of it, gets to know it all, surely he IS satisfying his wish to retrieve THAT VERY PIECE, to hear THOSE VERY MELODIES, is he not? We would say no, unfortunately. Doug cannot tell which ballet it was, he has no sense of joy at finding the precise long-lost piece. It is true that he has a NEW pleasure, but the old pleasure is never repeated. He actually hears the piece again, but without the proper intensional pointer to it. Oh, the pangs of intensionality!

CONCLUSION

Well, this about winds up our wild and woolly trip through the land of in- and ex- tensions. We have enjoyed rambling back and forth from rather simple and sedate ideas about frames and slots to absolutely wild-eyed speculative philosophical meditations. Any system which could handle all the issues we have raised here would have to be almost unimaginably fluid -- we say this, but then we remember that people are walking instantiations of such systems! It gives us tremendous respect for the subtlety of human minds to think of their awesome ability to recognize, classify, abstract, transfer, compare, distinguish, refine, reorganize. We hope, in this paper, to have surveyed some territory in which more detailed explorations will soon be made.

REFERENCES

[Anderson 78]
Anderson, John R. "The Processing of Referring Expressions within a Semantic Network", in TINLAP-2 ("Theoretical Issues in Natural Language Processing-2". New York: ACM, 1978.

[Brachman 77]
Brachman, Ronald J. "What's in a Concept: Structural Foundations for Semantic Networks". International Journal of Man-Machine Studies (1977) 9, 127-152.

[Brachman 78]
Brachman, Ronald J. "A Structural Paradigm for Representing Knowledge". Bolt Beranek and Newman Technical Report No. 3605. Cambridge, Mass., 1978.

[Brachman 79a]
Brachman, Ronald J. "An Introduction to KL-ONE", in "Research in Natural Language Understanding: Annual Report". Bolt Beranek and Newman Technical Report No. 4274. Cambridge, Mass., 1979.

[Brachman 79b]
Brachman, Ronald J. "On the Epistemological Status of Semantic Networks", in "Associative Networks" (Nicholas V. Findler, ed.). New York: Academic Press, 1979.

[Creary 79]
Creary, Lewis G. "Propositional Attitudes: Fregean Representation and Simulative Reasoning", in "Proceedings of the Sixth International Joint Conference on Artificial Intelligence". Stanford University Computer Science Department, 1979.

[Dennett 80]
Dennett, Daniel C. "Beyond Belief". (To be published)

[Gentner 80]
Gentner, Dedre. Personal communication.

[Hofstadter 79]
Hofstadter, Douglas R. "Godel, Escher, Bach: an Eternal Golden Braid". New York: Basic Books, Inc., 1979.

[Hofstadter-Dennett 81]
Hofstadter, Douglas R. and Daniel C. Dennett. "Soul Searching". (To be published)

[Kripke 1980]
Kripke, Saul A. "Naming and Necessity". Cambridge, Mass.: Harvard University Press, 1980.

[Minsky 75]
Minsky, Marvin. "A Framework for Representing Knowledge", in "The Psychology of Computer Vision" (P.H. Winston, ed.). New York: McGraw-Hill, 1975.

[Schank 79]
Schank, Roger C. "Reminding and Memory Organization: An Introduction to MOPs". Yale University Computer Science Department Research Report No. 170. New Haven: 1979.

[Sussman 75]
Sussman, Gerald J. "A Computer Model of Skill Acquisition". New York: American Elsevier, 1975.

[Winograd 72]
Winograd, Terry A. "Understanding Natural Language". New York: Academic Press, 1972.