THE LOGIC OF PLAUSIBLE REASONING:
A CORE THEORY

by

A. Collins
R. S. Michalski

THE LOGIC OF PLAUSIBLE REASONING:
A CORE THEORY

Allan Collins
Bolt Beranek and Newman Inc.
Cambridge, MA 02238

Ryszard Michalski
Computer Science Department
George Mason University
Fairfax, VA 22030

ABSTRACT

The paper presents a core theory of human plausible reasoning, based on analysis of people's answers to everyday questions about the world. The theory consists of three parts:

1. a formal representation of plausible inference patterns; such as deductions, inductions, and analogies, that are frequently employed in answering everyday questions.

2. a set of parameters, such as conditional likelihood, typicality, and similarity, that affect the certainty of people's answers to such questions, and

3. a system relating the different plausible inference patterns and the different certainty parameters.

This is one of the first attempts to construct a formal theory that addresses both the semantic and parametric aspects of the kind of everyday reasoning that pervades all of human discourse.

1. BACKGROUND FOR THE THEORY

The goal of our research on plausible reasoning is to develop a formal system based on Michalski s (1980, 1983) variable-valued logic calculus that characterizes different patterns of plausible inference humans use in reasoning about the world (Polya, 1958; Collins, 1978a). Our work attempts to formalize the plausible inferences that frequently occur in people's responses to questions for which they do not have ready answers (Carbonell & Collins, 1973; Collins, 1978a,b; Collins, Warnock, Aiello, & Miller, 1975). In this sense it is a major departure from formal logic and various non-classical logics: e.g., fuzzy logic (Zadeh 1965), multiple-valued logic (Lukasiewicz 1967), Dempster-Shafer logic (Shafer, 1976), intuitionist logic (Martin-Lof, 1982) variable-precision logic (Michalski & Winston 1986), probabilistic logic (Nilsson 1986), belief networks (Pearl 1986), and default logic (Reiter 1980, Yager 1987), Being descriptively based, the theory includes a variety of inference patterns that do not occur in formal logic-based theories. The central goals of the theory are to discover recurring general patterns of human plausible inference and to determine parameters affecting the certainty of these inferences. Unlike other theories of plausible reasoning, the theory combines semantic aspects with parametric aspects captured by numeric or symbolic estimates of certainty.

In order to analyze human plausible reasoning. Collins (1978b) collected a large number of people's answers to everyday questions, some from teaching dialogues and some from asking difficult questions to four subjects. These answers have the following characteristics:

1. There are usually several different inference patterns used to answer any question.

2. The same inference patterns recur in many different answers.

3. People weigh different evidence that bears on their conclusion.

4. People are more or less certain about their conclusion depending on the certainty of their information (either from some outside source or from memory), the certainty of the inference patterns and associated parameters used, and on whether different patterns lead to the same or opposite conclusions.

The analysis of the answers attempts to account for the reasoning and the conclusions drawn in terms of a taxonomy of plausible inference patterns. As will be evident, this is an inferential analysis. To use Chomsky's (1965) felicitous terms, we are trying to construct a deep structure theory from the surface structure traces of the reasoning process.

In our development of the theory to date we have not tried to characterize all the different types of plausible inferences that occur in the protocols. In particular we have not formalized the spatial, temporal, and meta-knowledge inferences often seen in protocols (Collins 1978a). This project presents a core system centered around the plausible deductions, analogies, and inductions, seen most frequently in the protocols, but we expect there are other forms of these inferences that will need to be added to the core theory. In future work we plan to extend this core system to encompass the other patterns of inference, such as spatial, temporal, and meta-knowledge inferences (Collins, 1978 a,b).

We will illustrate some of the characteristics of people's answers, as well as some of the inference patterns formulated in the theory with several transcripts. The first transcript comes from a teaching dialogue on South American geography (Carbonell & Collins, 1973) (T stands for teacher and S for student):

T. There is some jungle in here (points to Venezuela) but this breaks into a savanna around the Orinoco (points to the Llanos in Venezuela and Colombia).

S. Oh right, is that where they grow the coffee up there?

T. I don't think that the savanna is used for growing coffee. The trouble is the savanna has a rainy season and you can't count on rain in general. But I don't know. This area around Sao Paulo (in Brazil) is coffee region, and it is sort of getting into the savanna region there.

In the protocol the teacher went through the following reasoning. Initially, the

3

teacher made a hedged "no" response to the question for two reasons. First, the teacher knew that coffee growing depends on a number of factors (e.g., rainfall, temperature, soil, and terrain), and that savannas do not have the correct value for growing coffee on at least one of those factors (i.e., reliable rainfall). In the theory this is an instance of the inference pattern called a *derivation from a mutual implication*[1], in particular the implication that coffee growing depends on reliable rainfall. Second, the teacher did not know that the Llanos was used for growing coffee, which he implicitly took as evidence against its being a coffee region. The inference takes the form "I would know the Llanos produces coffee if it did, and I don't know it, so probably it does not." This is called a *lack-of-knowledge inference* (Collins et al., 1975; Gentner & Collins, 1982). This inference pattern is based on knowledge about one's own knowledge and hence is a meta-knowledge inference.

Then the teacher backed off his initial negative response, because he found positive evidence. In particular, he thought the Brazilian savanna might overlap the coffee growing region in Brazil around Sao Paulo, and therefore might produce coffee. If the Brazilian savanna produces coffee, then by functional analogy (called a *similarity transform* in our theory) the Llanos might. Hence, the teacher ended up saying "I don't know," even though his original conclusion was correct.

The teacher's answer exhibits a number of the important aspects of human plausible reasoning. In general, a number of inference patterns are used together to derive an answer. Some of these are inference chains where the premise of one inference draws on the conclusion of another inference. In other cases the inference patterns are triggered by independent sources of evidence. When there are different sources of evidence, the subject weighs them together to determine a conclusion and the strength of belief in it. This weighing of evidence parallels the theory of endorsements espoused by Cohen (Cohen 1985; Cohen & Grinberg, 1983).

It is also apparent in this protocol how different pieces of information are found over time. What appears to happen is that the subject launches a search for information starting with the words in the question (Quillian, 1968; Collins & Loftus,

---

[1] This and other technical terms introduced in italics in the paper are defined and exemplified in a Glossary at the end of the paper.

1975). As pieces of information are found, they trigger particular inferences. Which inference pattern is applied is determined by the relation between the information found and the question asked. For the question about growing coffee in the Llanos, if the respondent knew that savannas are in general good for growing coffee, that would trigger a deductive inference. If the respondent knew of a similar savanna somewhere that produced coffee, that would trigger an analogical inference. In the protocol, the more accessible information about the unreliable rainfall in savannas was found before the less accessible information about the coffee growing region in Brazil and its relation to the Brazilian savanna. The search for information is such that the most accessible information is found first, as by a marker passing or spreading activation algorithm (Charniak, 1982; Quillian, 1968).

The next protocol illustrates a plausible deduction. (Q stands for questioner and R for respondent).

Q. Is Uruguay in the Andes Mountains?

R. I get mixed up on a lot of South American countries (pause). I'm not even sure. I forget where Uruguay is in South America. It's a good guess to say that it's in the Andes Mountains because a lot of the countries are.

The respondent knew that the Andes are in most South American countries (7 out of 9 of the Spanish speaking countries). Since Uruguay is a fairly typical South American country, he guesses that the Andes may be there too. He is wrong, but the conclusion was quite plausible. This kind of plausible deduction is called a *specialization transform* in the theory, based on the fact that Uruguay is a specialization of a South American country. This example illustrates two of the certainty parameters associated with it . *frequency* (he knows the Andes are in most countries), and *typicality* (Uruguay is a typical South American country).

The third protocol illustrates the other kind of plausible deduction in the theory, called a *derivation from mutual implication* (in particular, rice growing implies warm weather, flat terrain, and fresh water).

Q   Do you think they might grow rice in Florida?

R. Yeah. I guess they could, if there were an adequate fresh water supply.
    Certainly a nice, big, warm, flat area.

The respondent knew that whether a place can grow rice depends on a number of factors. He also knew that Florida had the correct values on at least two of these factors (warm temperatures and flat terrain). He therefore inferred that Florida could grow rice if it had the correct value on the other factor he thought of (i.e., adequate fresh water). He may or may not have been aware that rice growing also depends on fertile soil, but he did not mention it here. Florida in fact does not produce rice in any substantial amount, probably because the soil is not adequate. This protocol shows how people make plausible inferences based on their approximate knowledge about what depends on what, and how the certainty of such inferences is a function of the degree of dependency between the variable in question (rice) and the known variables (i.e. terrain, climate, water).

The fourth protocol from a teaching dialogue illustrates two inferences in the core theory, a *similarity transform* and a *dissimilarity transform*:

S. Is the Chaco the cattle country?   I know the cattle country is down there
    (referring to Argentina).

T. I think it's more sheep country.   It's like western Texas, so in some sense I
    guess it's cattle country.   The cattle were originally in the Pampas, but not
    so much anymore.

As in the first protocol, the respondent is making a number of plausible inferences in answering this question, some of which lead to different conclusions. First, he thinks that the Chaco is used for sheep raising, but there is some uncertainty about the information retrieved, which leads to a hedged response. This supports a *dissimilarity transform* and an implicit *lack-of-knowledge inference* (a

6

meta–knowledge inference). The *dissimilarity transform* is based on the view that sheep country is distinct from cattle country, presumably in terms of its climate or vegetation, so that if the Chaco is sheep country it is not likely to be cattle country. The *lack–of–knowledge inference* takes the form "I don't know that it's cattle country, and I would know if it were (e.g., I know about sheep), so it probably is not cattle country." But then the teacher noted a similarity between the Chaco and western Texas, presumably in terms of the functional determinants of cattle raising (e.g., climate, vegetation, terrain). Because Western Texas is cattle country, this led him to a very hedged affirmative response, based on a *similarity transform.* Finally the teacher alluded to the fact that the Pampas is the place in Argentina known for cattle, and the place the student most likely was thinking of. This argues against the Chaco having cattle based on another meta–knowledge inference, a *functional alternative inference* (Collins, 1978b; Pearl, 1987): "The Pampas is an Argentinan plain and the Pampas has cattle, so the fact that there are cattle in an Argentinan plain cannot be taken as evidence for cattle in the Chaco." In answering this question, then, two patterns of plausible inference led to a negative conclusion and one to a positive conclusion.

The fifth protocol again illustrates both a *similarity* and a *dissimilarity transform,* and more importantly, the distinction between inferences based on overall similarity and those based on similarity with respect to the functional determinants of the property in question.

Q. Can a goose quack?

R. No a goose – well, its like a duck, but its not a duck.

   It can honk, but to say it can quack. No, I think its

   vocal cords are built differently. They have a beak

   and everything, but no, it can't quack.

The *similarity transform* shows up in the phrases, "it's like a duck" and "They have a beak and everything" as well as in the initial uncertainty about the negative conclusion. It takes the form, "A duck quacks and a goose is like a duck with respect

to most features, so maybe a goose quacks". The certainty of the inference depends on the degree of similarity between ducks and geese.

But then two lines of negative inference led the respondent to a negative conclusion. First there is a *lack-of-knowledge inference* implicit in the statement "It can honk, but to say it can quack." She knew about geese honking but not about their quacking. Therefore, she thought she would know about geese quacking, if in fact they did quack.

The second line of negative inference (apparently found after she started answering) is the *dissimilarity transform* evident when she says, "I think its vocal cords are built differently". The *dissimilarity transform* takes the form "Ducks quack, geese are dissimilar to ducks with respect to vocal cords, and vocal cords determine the sound an animal makes, so probably geese do not quack". This inference was enough to lead her to a strong "no". Of course she knew nothing about the vocal cords of ducks and geese, because they don't have any. She was probably thinking of the difference in the length of their necks. Our own hypothesis is that longer necks resonate at lower frequencies and hence honking can be thought of as deep quacking.

These five examples illustrate a number of aspects of human plausible reasoning as it occurs in common discourse. They show how people bring different pieces of knowledge to bear on a question and how these pieces sometimes lead to the same conclusion and sometimes to different conclusions. Often knowledge is found after the respondent has started answering, so that the certainty of the answer seems to change in midstream. The examples also show how people's approximate functional knowledge of what depends on what often comes to play in different inferences such as deductions and analogies. Therefore these dependencies are a central part of the core theory we have developed. We will return to these examples to illustrate how the formal rules we have developed can be used to characterize different plausible inferences seen in these examples.

We have collected many such protocols (Collins 1978b) and these same patterns (as well as others) recur again and again in many different content domains and contexts. Any theory that is to account for such data will have to characterize these systematic patterns and the way that functional dependencies (e.g. coffee growing

depends on reliable rainfall) interpenetrate these patterns of inference. The theory outlined in the rest of the paper is the simplest theory that we have been able to construct to do that job.

We should emphasize also that the scope of the theory is the kind of domain-independent, weak inferences (Newell 1980) akin to the syllogistic forms in logic. The core theory attempts to specify the generalizations of syllogistic forms that reflect the way people actually reason, not how they should reason. This scope leaves out two kinds of plausible reasoning seen frequently in people's answers to questions: 1) domain specific reasoning (e.g. "the language of Mexico is Mexican," which employs a special rule for forming language names), and 2) generalized weak methods that involve active search for information, such as means-ends analysis (Newell & Simon, 1972) and proof by cases (e.g., to estimate how many Catholics there are in the world, many people will consider different countries or continents and estimate how many in each).

Johnson-Laird (1980, 1983) has argued that the best account for human reasoning is not in terms of systematic rules or inference patterns, but rather in terms of the manipulation of mental models. While we agree that people manipulate mental models in their reasoning (Collins, 1985; Collins & Gentner, 1982, 1983, 1987; Stevens & Collins, 1980), their use of mental models is orthogonal to the systematic patterns described in this paper. In particular, the protocols we have collected often involve picturing different situations (e.g. a mental map of South America, images of savannas, or an advertisement showing Juan Valdez on his coffee plantation in Colombia). These images can be taken as evidence for the manipulation of mental models in Johnson-Laird's terms. But overlaying this manipulation of mental models are the systematic patterns in which they are deployed to support one's conclusions (c.f. Rips 1986). So while mental models may be part of the story of plausible reasoning, there is another critical part which the theory we propose addresses.

The theory does not address the issue of whether people make systematic errors in their reasoning, as the psychological literature on decision making (Kahneman, Slovic, and Tversky, 1982) attempts to document. This issue does not arise in the theory because we are developing a formalism for representing the kinds of inferences people make and the parameters that affect their certainty, rather than a theory about how people make particular inferences. People may systematically ignore some

kinds of information or undervalue particular certainty parameters – we have not attempted to determine whether they do or not. Instead we have tried to represent all the kinds of reasoning patterns and the kinds of certainty parameters that appear in the protocols we have analyzed (Collins, 1978 a, b). In this regard it is worth pointing out that certain fallacies in logic, such as affirming the consequent (Haviland, 1974), become plausible inference patterns in the theory.[2]

The theory was developed to account for protocols where a question drives the search for relevant information – in Artificial Intelligence this is called backward inferencing. One question that might arise is whether the theory applies to forward inferencing: when a person finds out some fact such as that they grow rice in Java, does she draw inferences about places that might grow rice (e.g. Sumatra, Borneo, the Philippines, or even Madagascar, the Congo, and Brazil) or about what conditions lead to rice growing (e.g. a tropical climate, an oriental location, an island climate, having a lot of people to feed, etc.). One danger in forward inferencing is that there are so many possible inferences, it can go on forever – if one decides that islands can grow rice, one can carry this to Iceland and Greenland and then wonder about Antarctica and Australia, or even Africa. In general people probably do not do much forward inferencing, except as Schank (1986) suggests when they ask themselves questions in order to explain and generalize their experiences. In any case, people do some forward inferencing and our guess is that the same patterns occur. But they do not carry it very far because the certainty of the inferences quickly falls below some threshold of plausibility. There are just no long chains of inference in people's plausible reasoning, unlike logical or mathematical proofs.

There is a high payoff from trying to formalize the patterns of human plausible reasoning in a system. The system helps to identify parallels between apparently different inference types (e.g. deductions, analogies and inductions). Furthermore, it makes it possible to see the systematic patterns in which different certainty factors in the psychological literature (e.g. typicality, similarity, frequency, dominance) affect related inference types. The payoff is similar to what happened when Mendeleev

---

[2]As will be seen, dependencies and implications are bidirectional in the theory and so derivations from them, such as affirming the consequent, are plausible but not certain inferences. The same point is made by Polya (1968).

discovered the periodic table establishing the regularities between different chemical elements. then it was possible to see which elements were missing, predict how new elements might behave. and begin the search for why the systematic patterns in the table arose at some deeper level.

2. ASSUMPTIONS UNDERLYING THE THEORY

The theory assumes that a large part of human knowledge is represented in *dynamic hierarchies*, that are always being updated, modified or expanded. In the *core* theory described here we distinguish between two basic kinds of hierarchies, *type*- and *part*-hierarchies (Collins and Quillian, 1972). A *type*-hierarchy (also called an *abstraction* or *is—a* hierarchy) is organized by the *type* relation holding between connected nodes, or more precisely, between concepts represented by the nodes. A *part*-hierarchy is organized by the *part—of* relation holding between connected nodes. Any given node may be a member of more than one hierarchy. Each such hierarchy characterizes the node from a different viewpoint. There are other kinds of hierarchies (e.g. kinship hierarchies) that govern human inferences, but they play a minor role as compared to type- and part-hierarchies.

Nodes of a hierarchy may represent classes (e.g., flowers), individuals (e.g., a specific flower) or manifestations of individuals (e.g., a specific flower at a given moment). For the purpose of the theory, they are treated alike, though it may be necessary in future refinements of the theory to treat manifestations, individuals, and classes differently.

Figure 1 shows examples of type- and part-hierarchies. In the first four examples (a,b,c,d), the Llanos is viewed from four different perspectives. These perspectives are organizing principles of the hierarchies (Bobrow and Winograd, 1977). The type-hierarchy in Figure 1a is organized according to the type of terrain. The type of terrain can be mountainous, plateau, hilly, or plain, etc. The Llanos is characterized as a type of plain, like the Chaco. The type-hierarchy in Figure 1b is organized according to the geographical land type. It characterizes the Llanos as a type of savanna, which is one of the major land types that geographers divide the world into, including rain forests, deserts, steppes, Mediterranean climates, mid-latitude forests, etc. The part-hierarchy in Figure 1c is organized according to regions in South America. the Andes, Amazon Jungle, Llanos, Guiana Highlands, and their subregions in different countries. The part-hierarchy in Figure 1d represents South America broken down into countries and the subregions within each.

12

The other three examples in Figure 1 are designed to illustrate how different kinds of information are represented in hierarchies. Among colors there are green and red. Among reds there are scarlet and burgundy, and among scarlets there are bright scarlet and perhaps dull scarlet, etc. Color is a one-place descriptor applying to objects, but feeling emotion is a two place descriptor where X (a person) feels the emotion toward Y (any concept). In the emotion hierarchy there are many types of emotions, among them love and hate, and there are different kinds of love, such as romance, affection, motherly love, etc. In the weight hierarchy there are different kinds of weight, such as human weight which in turn might be divided into birth weight and adult weight. For birth weight one might think of 1 lb. as a minimum, 15 lbs as a maximum, and 7 lbs as the norm. For the purposes of the theory these can be thought of as different values of birth weight, just as red and green are different values of color.

Node A in any hierarchy can be a *descriptor* of node B in another hierarchy, i.e., A can be used to characterize node B. We write such a relation as a *term*

$$A(B)$$

For example, the node "color" in hierarchy 1e of Figure 1 applies as a descriptor to the node "eyes" in a hierarchy of body parts. This is denoted as "color(eyes)." The node "eyes" can in turn be applied as descriptor to the node, John, in some hierarchy describing people. To express both relations we would write.

$$color(eyes(John))$$

A term A(B) can take values (called *referents* in the theory) only from the set of subnodes of A, i.e., the descendants of the node A in the hierarchy. Applying a descriptor to an *argument* (a node or a sequence of nodes) produces a specific value characterizing the argument. This implies that only non-terminal nodes of a hierarchy can be descriptors. For example, to state that the color of the eyes of
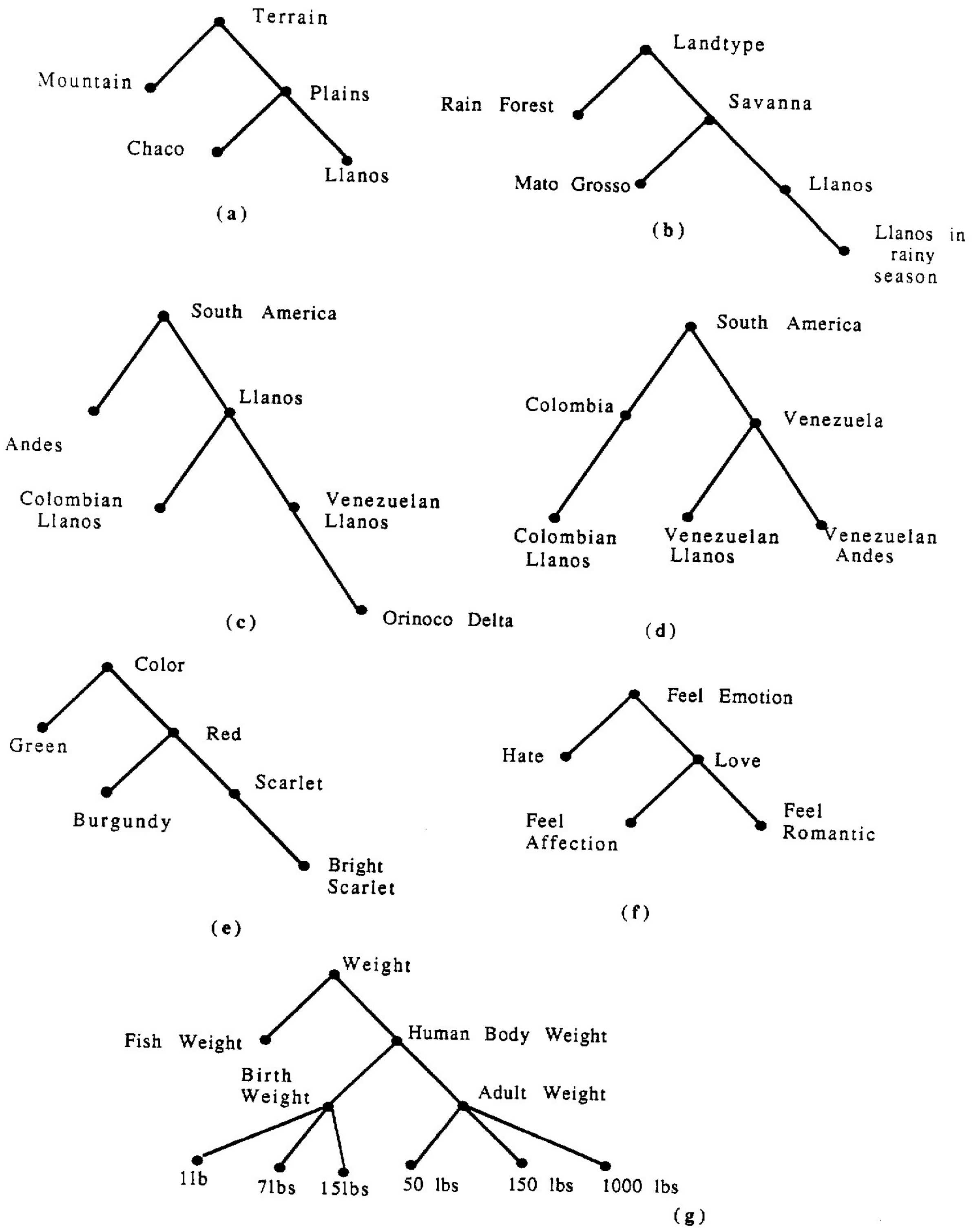
Figure 1. Examples of hierarchies.

John is blue, a path would be created that links John, color and blue as shown in Figure 2. To express this formally, we write:

$$color(eyes(John))=blue$$

In the theory such an expression is called a *statement*. Statements are recordings of information within the hierarchies. They are paths connecting the nodes of two or more hierarchies that represent beliefs about the world. Figure 2 shows examples of statements representing the beliefs that there are daffodils and roses in England, and that John's eyes are blue. The statements can have annotations describing their origin, their frequency of use, the certainty of belief in their correctness, and other information. The links denoting the type and part relation in generalization hierarchies can also be viewed as denoting statements, but for the purposes of the theory we will distinguish them from other statements. The knowledge organization described above includes various elements of semantic network structure (Carbonell & Collins, 1973; Collins & Quillian, 1972; Quillian, 1968; Woods, 1975) and frame structure (Bobrow & Winograd, 1977; Brachman & Schmolze, 1985; Minsky, 1975; Schank & Abelson, 1977; Winograd, 1975).


Insert Figure 2 here

Figure 3 illustrates the fact that the hierarchies are partial orderings, and can be differentiated or collapsed as appropriate for the purpose of drawing plausible inferences. At a fairly early age children think of animals as coming in different types: dogs, cats, fish, birds, etc. They don't differentiate them much more than that. When they get to school they may learn there are different basic types of animals, such as fish, birds, reptiles, mammals, and amphibians, and that dogs and cats are types of mammals. Still later in biology this hierarchy might be differentiated much more finely as in Figure 3c. For the purposes of the theory, such hierarchies may be thought to coexist, and plausible inferences can be made in any of them.
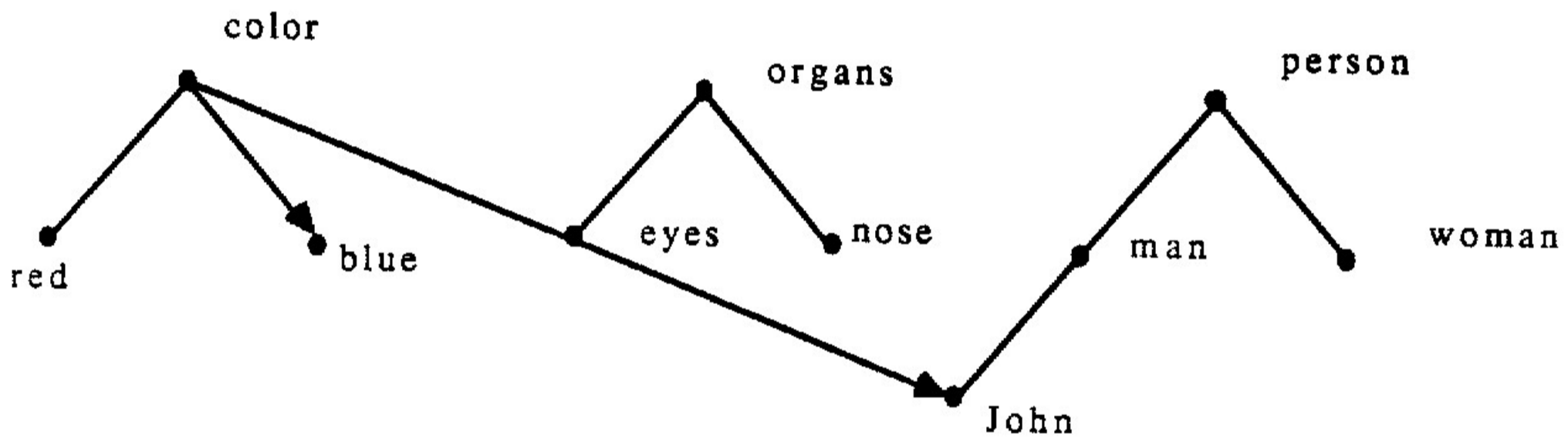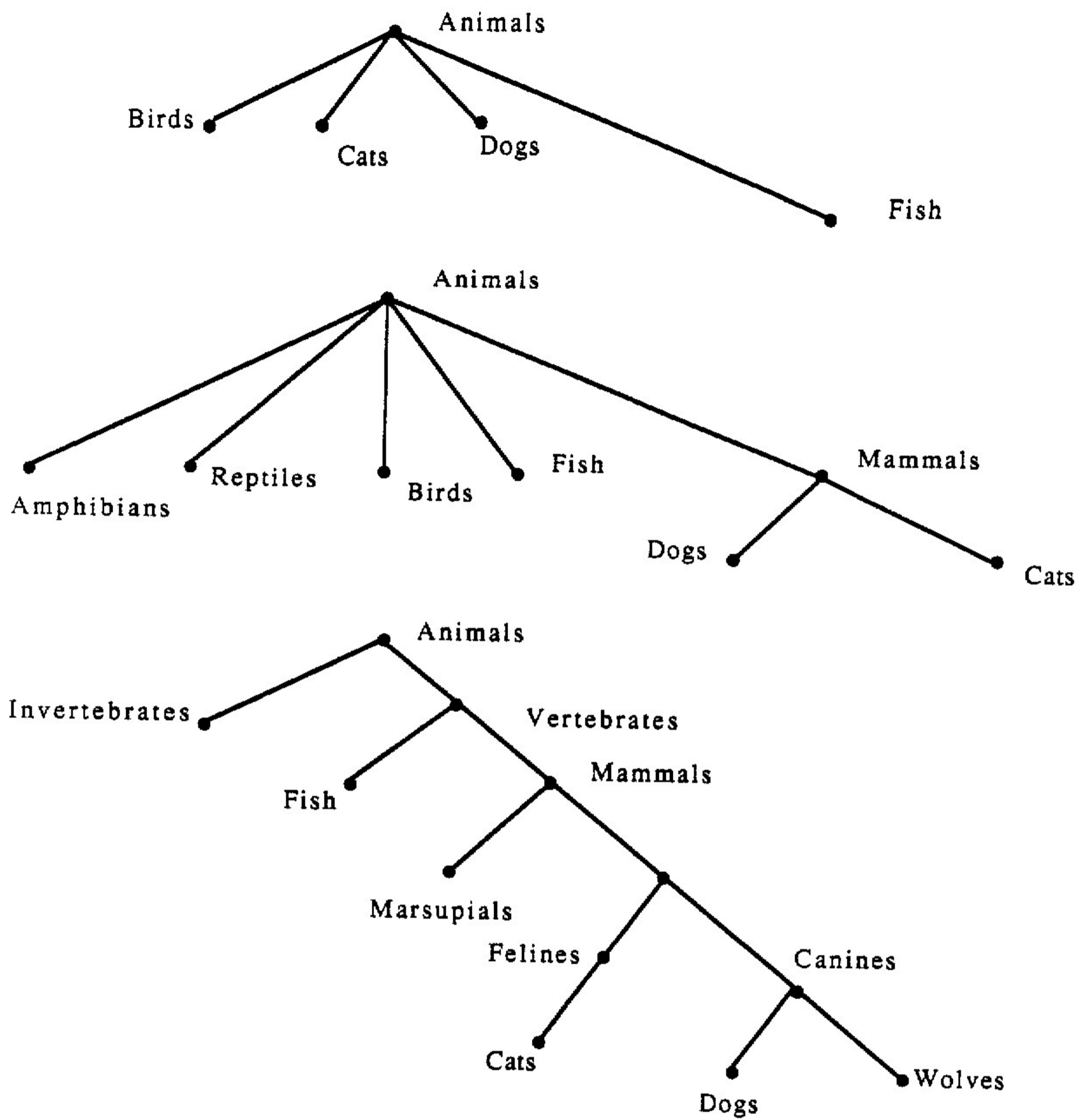

Insert Figure 3 here

---
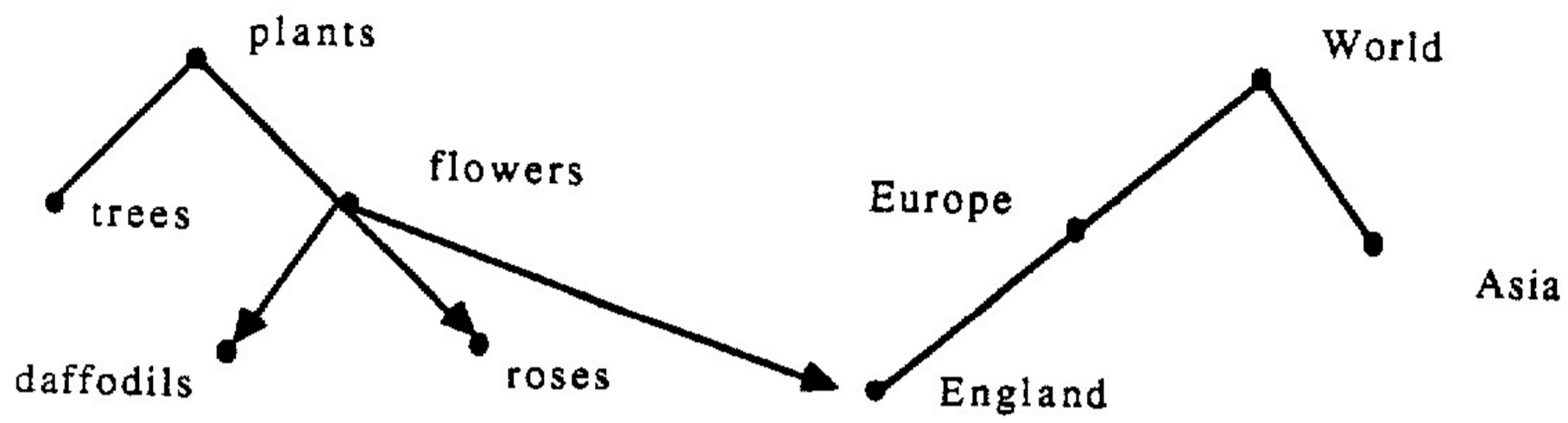
flowers (England) = {daffodils, roses. . .}

color (eyes(John)) = blue

Figure 2. Examples of two traces on statements

Figure 3: Differentiation of Hierarchies

Table 1 shows hypothetical concept structures for a few concepts in someone's memory (Collins & Quillian, 1972; Collins et al, 1975). These examples are not meant to provide a detailed analysis of how concepts are represented, but rather to illustrate how the statements shown in later examples can be constructed from a memory structure. In the example, type and part relations form the basis for hierarchical structures such as those shown in Figures 1 and 3. Flowers are represented as a type of plant coming in at least four varieties (i.e. roses, etc.), having various parts, various colors, and growing in all countries. Each descriptor (i.e. type/of, types, parts, color, countries) might be further specified as to how it relates to the concept flower (e.g., type/of is a biological class, colors are surface features of the petals, countries are places where flowers are grown, etc). The concept description for daffodils, which are a particular type of flower, provides further specification for each of the variables in the concept of flowers. That is, they have petals and a stem, they come in yellow and perhaps other colors, and they are grown in at least England and the United States. The concept red is shown to illustrate how a color concept points back to various objects which it describes. Finally let us stress that we have not concerned ourselves with exactly how concepts are represented, but rather we have assumed they are represented in a structure similar to these examples.

Insert Table 1 here

Any node in a hierarchy can potentially be a *descriptor* for a node in another hierarchy. For example, if flower is in a hierarchy of things and England is in a hierarchy of places, flower-type might be a descriptor for England. This produces a *statement* of the form:

(1) flower-type(England)={daffodils, roses,....}

In (1) flower-type is a *descriptor*, England is an *argument*, flower-type(England) is a *term*, and daffodils and roses are *referents* for the term. The brackets and dots indicate that daffodils and roses are not assumed to be a complete set, although the person may not know other flowers of England. Any descriptor, as a node in a hierarchy, can be further differentiated. For example, flowers can be differentiated between naturally-growing flowers vs. flowers grown in greenhouses, or between flowers sold vs. flowers grown, etc. People make finer or less fine discriminations

15

Table 1

Hypothetical Frames in a Person s Memory

flower

    type/of = plant

    types ={rose, daffodil, peony, bougainvillea ...}

    parts ={petals, stem ...}

    colors ={pink, yellow, white, red ...}

    countries ={all countries}

daffodil

    type/of = flower

    parts ={petals, stem ...}

    colors ={yellow ...}

    countries ={England, United States ...}

red

    type/of = color

    types ={scarlet, burgundy ...}

    flowers ={roses, tulips ...}

    vehicles ={fire engines, London buses ...}

depending on their knowledge and purposes, and a theory of plausible reasoning must accommodate these different degrees of discrimination.

Whether a particular descriptor applies to any argument depends on what knowledge the person has. For example, it is not clear what color-type (England) might mean because one probably doesn't have knowledge in one's data base about the color of England (though one might interpret the term as the color of the overall appearance of the country, e.g., Ireland looks green).

Examples (2) to (6) below illustrate how different descriptors apply to different concepts:

(2) England-part(daffodils)={Southern England...}

(3) daffodil-part(England)={petals, stem...}

(4) country-type(daffodils)={temperate countries...}

(5) daffodil-type(England)={yellow daffodils...}

(6) England-type(daffodils)={England in the spring}

Examples (2) and (3) illustrate statements based on part hierarchies. In (2) the descriptor selects the part of England where daffodils occur. In (3) the descriptor selects the parts of daffodils that occur in England; presumably daffodil parts in England are the same as daffodil parts anywhere in the world (though perhaps Martian daffodils are quite different). In (4) country-type applied to daffodils selects the types of countries that have daffodils (i.e., temperate countries). Statement (4) could have specified the particular countries (e.g., England, France) that have daffodils, since hierarchies can be collapsed as long as a partial order is maintained. In (5) daffodil-type applied to England selects the different daffodil types found in England, of which only one type is stored (i.e., yellow daffodils), though there may be others. In (6) we show that when you take an instance like England and look at its subtypes you get a manifestation, in this case the manifestation that has daffodils. These examples show how different terms are evaluated within the theory.

We have discussed the most important assumptions we are making about how

human memory is organized and accessed for the purposes of making plausible inferences. Further descriptions of our underlying assumptions about human memory are given in earlier papers (Carbonell & Collins, 1973; Collins & Loftus, 1975; Collins & Quillian, 1972; Collins, Warnock, Aiello & Miller, 1975).

3. PRIMITIVES IN THE CORE SYSTEM

In the core system we have developed there is a set of primitives and a set of basic inference rules. In this section we describe the primitives in the system, consisting of basic expressions, operators, and certainty parameters.

Table 2 shows the basic elements in the core system. *Arguments* can be any node in a hierarchy, or a function of one or more nodes such as Fido's master or the flag of England. *Descriptors* apply to arguments, and together they form a *term*, such as breed (Fido). The potential *referent* for a term is the set of nodes in the hierarchy under the descriptor node: it can be either a definite set of values such as collie, or brown and white, or an indefinite set of values such as brown plus other colors (or possibly no other colors). Indefinite sets are represented by brackets and dots (e.g., {brown...}).

Insert Table 2 here

Statements consist of a term on the left of a relational operator (usually an equals sign) and a referent on the right, together with a set of certainty parameters. Expressions (1) through (6) above were all statements, without the certainty parameters specified. The certainty parameters can be thought of as approximate numbers ranging between 0 and 1, but we have represented them as verbal descriptions. In the example of a statement in Table 2, γ refers to how certain one is the statement is true, and ϕ to the belief about the frequency that if something is a bird it can fly (p(flying/bird)). These certainty parameters are all listed in Table 4, to be discussed later.

The last two types of expressions in Table 2 represent functional dependencies between different variables. *Mutual dependencies between terms* represent the functional relationship between two terms, such as between the average temperature of

18

Table 2

Elements of Expressions

arguments
a_1, a_2, $f(a_1)$
   e.g., Fido, collie, fido's master

descriptors d_1, d_2
   e.g., breed, color

terms $d_1(a_1)$, $d_2(a_2)$, $d_2(d_1 (a_1))$
   e.g., breed (Fido), color (collie), color (breed (Fido))

referents r_1, r_2 r_3, $\{r_2 ...\}$
   e.g., collie, brown and white, brown plus other colors

statements $d_1(a_1)=r_1$:γ, ϕ
   e.g., means-of-locomotion (bird)={flying...}: certain. high
      frequency (translation: I am certain almost all birds fly)

dependencies between terms $d_1(a_1)<--->d_2(f(a_1))$:α,
β, γ
  e.g., latitude (place) <----> average temperature (place).
     moderate. moderate. certain (translation: I am certain
     that latitude constrains average temperature with moderate
     reliability. and that average temperature constrains latitude
     with moderate reliability)

implications between statements
$d_1(a_1)=r_1 <===>d_2(f(a_1))=r_2$: α, β,
γ
  e.g., grain (place)={rice...} <===> rainfall (place)=heavy.
     high, low, certain (translation. I am certain that if
     a place produces rice. it implies the place has heavy
     rainfall with high reliability. but that if a place
     has heavy rainfall it only implies the place produces
     rice with low reliability)

a place and the latitude of the place.[3] The dependency can be annotated to different degrees: it can be unmarked meaning there exists some functional relation between the two, it can be marked with + or − indicating a monotonic increasing or decreasing relation, or it can be further specified to any degree (e.g., a V−shaped function with 3 values specified). For example, if one thinks that average temperature of a place in January varies between about 85° at the equator and −30° at the North Pole and + 30° at the South Pole, this relation can be represented as a V−shaped function with values (−90°, 30°), (0°, 85°) and (90°, −30°), where the first number is latitude and the second temperature. The conditional likelihood parameters (α and β) specify the degree of constraint in the dependency from latitude to temperature and from temperature to latitude, respectively. In the latitude−temperature example the degree of constraint is moderate in both directions, as is discussed later.

Mutual implications between statements relate particular values of functions such as the latitude−temperature function above (e.g., latitude (place) = equator <=> average temperature (place) = hot). The example shown in the table relates the grain of a place being rice to the rainfall of the place being heavy (e.g., >40 in/year). Knowing a place produces rice predicts that it will have heavy rainfall quite strongly, so that α is high (though there are exceptions like Egypt where rice is grown by irrigation). However the fact that the rainfall of a place is heavy (e.g., Oregon) only weakly predicts that rice is grown, so β is low. In general mutual implications between statements will be asymmetric in this way.

Table 3 illustrates the four relations in the core system and the kinds of statements they occur in. The generalization (GEN) and specialization (SPEC) relations go up and down in a hierarchy, while the similarity (SIM) and dissimilarity (DIS) relations go between any two comparable nodes in a hierarchy. Associated with the

---

[3]Dependencies between terms play essentially the same role in this theory as *determinations* play in the work of Davies and Russell (1986). Determinations have the form that one variable determines the truth of another variable, whereas dependencies are bidirectional, where each variable constrains the other variable to some degree specified by certainty parameters α and β. Determinations are used in their theory to determine the relevance of variables over which analogies are made, whereas dependencies in our theory more generally constrain a larger class of inferences including basic analogies (called similarity transforms in our theory).

GEN and SPEC relations there is a typicality[4] parameter τ (Rosch, 1975; Smith & Medin, 1982), and with the SIM and DIS relations there is a similarity parameter σ. There is also a dominance parameter δ associated with GEN and SPEC statements that specifies what proportion of the superset, the subset actually comprises. Finally all the statements involving relations have a η parameter associated with them reflecting the certainty of belief that the statement is true.

Insert Table 3 here

Typicality and similarity are always computed in some context (CX) which is denoted by the CX variables. The first variable in the CX denotes a node in the argument hierarchy specifying the range of arguments over which typicality or similarity are computed. For GEN and SPEC this is always the superset specified in the statement: for chicken SPEC barnyard fowl, barnyard fowl is the superset over which typicality is computed. For SIM and DIS, however, it is the basic level category (Rosch 1975; Smith & Medin, 1982) to which the two arguments belong that is the basis for computing similarity. Hence the similarity of ducks and geese would normally be computed in the context of birds, which is their basic level category.

The second variable in the CX specifies the set of descriptors to be used in comparing the two nodes with respect to typicality or similarity. For example, one can evaluate how typical chickens are as birds with respect to their physical features, with respect to all their features, or with respect to some particular feature such as the cost of feeding them. Similarity and dissimilarity can also be computed with respect to different features. As we discussed with respect to the fifth protocol shown earlier, ducks and geese are quite similar when compared on all their features, but they are dissimilar in neck length (which is relevant to determining the sound they make). The procedure for computing typicality and similarity is described below.

Table 4 lists the certainty parameters we have identified so far that affect the certainty of different plausible inferences. These parameters do not yet have an

---

[4]Typicality corresponds roughly to representativeness in the work of Kahneman and Tversky (1972).

agreed computational definition, and so different computer models of the theory have implemented them in different forms. We will describe each of these parameters in terms of the examples given above. The description is meant to specify our best hypothesis about how people might compute these parameters.

Insert Table 4 here

The conditional likelihood (α and β) parameters can best be introduced in terms of the example. grain(place)={rice...}<===>rainfall(place)=heavy. As we said, α would be high in such a case if a person thinks that most places that grow rice have heavy rainfall (say greater than 40 inches per year), whereas β would be low if he or she thinks there are many places with heavy rainfall, that don't produce rice. We can construct a hypothetical contingency table that represents this view in terms of a small sample of places and the frequencies with which they have heavy rainfall and produce rice:

	Rice	No Rice	Total
Heavy Rainfall	8	8	16
No Heavy Rainfall	2	20	22
Total	10	28	38

Given this table α is simply the conditional probability that a rice-producing place has heavy rainfall, in this case 8 of 10 or .8, and β is the conditional probability that a place with heavy rainfall produces rice, in this case 8 of 16 or .5. We don't think that people actually construct such tables though they may consider a small number of cases in computing rough estimates of α and β, as they do in using the availability heuristic (Tversky & Kahneman, 1973). Basically our assumption is that people build up a rough intuition about how frequently one thing leads to (or predicts) another, and this is what is captured by the α and β parameters. By providing two parameters, the theory can encompass the kind of asymmetries found by Tversky and Kahneman (1980) in reasoning causally vs. diagnostically.

The α and β parameters for mutual dependencies can be constructed by an extension of the procedure for mutual implications. Suppose one considers the relationship of rainfall and grain growing as before, but instead as a mutual

Table 3

Relations*

Generalization   a' GEN a in CX(a',d(a')): γ, τ, δ

    e.g., bird GEN chicken in CX (birds, physical features(birds)):
        certain, atypical, low dominance (translation: I am
        certain chickens are birds, but they are atypical of
        birds in their physical features, and they are a low
        percentage of birds)

Specialization   a' SPEC a in CX(a,d(a)): γ, τ, δ
    e.g., chicken SPEC fowl in CX (fowl, food cost(fowl)):
        certain, typical, moderate dominance
        (translation: I am certain chicken are fowl
        and they are typical of fowl with respect
        to food costs, and they are a moderate percentage of
        barnyard fowl)

Similarity   a' SIM a in CX(A,d(A)): γ, σ

    e.g., ducks SIM geese in CX(birds, all features(birds)): certain,
        highly similar (translation: I am certain ducks are
        highly similar to geese with respect to all their features)

Dissimilarity   a' DIS a in CX(A,d(A)): γ, σ

    e.g., ducks DIS geese in CX(birds, neck length(birds)): certain,
        fairly dissimilar (translation: I am certain ducks are
        fairly dissimilar to geese with respect to neck length)

* A represents a superordinate of a and a'

dependency: i.e., grain (place) <--> rainfall (place). For simplicity we can present the same hypothetical table in revised form:

	Rice	Wheat	Corn	Total
Heavy Rainfall	8	6	2	16
Light Rainfall	2	14	6	22
Total	10	20	8	38

Then α reflects the degree to which you can predict whether a place has heavy or light rainfall, given the predominant grain grown in the place, which is quite high (i.e., the prediction is correct in 28 or 38 cases or .7 assuming an optimal guessing strategy). Similarly, β reflects the degree to which you can predict whether they grow rice, wheat, or corn, given the amount of rainfall (i.e., the prediction is correct in 22 of 38 cases or .6, assuming an optimal strategy of guessing wheat for light rainfall and rice for heavy rainfall). This example makes evident the fact that the α and β parameters reflect the way the dependency partitions the known cases in the world.

The γ parameter in Table 3 reflects the certainty or subjective likelihood with which a person believes any expression is true. γ can reflect different possible sources of uncertainty. One source arises when people retrieve a fact from memory and are uncertain whether they are making a memory confusion. Another basis for uncertainty arises when they doubt the source from which they got the information. Finally, if a piece of information derives from a plausible inference, there will be uncertainty as to whether the conclusion is correct, and this uncertainty will propagate to inferences dependent on it. All these sources of uncertainty are represented by the γ parameter.

Typicality (τ) and similarity (σ) both involve computing the coincidence of features.[5] In the case of typicality it is computed between a subset and its superset, and in the case of similarity it is computed between two subsets. We assume for these purposes that any set (or concept) is represented as a bundle of features (Collins & Quillian, 1972), and the τ and σ parameters are computed by comparing the two concepts with respect to those features specified by the descriptor variable in the

---

[5]Rips (in press) has evidence that people in fact compute typicality and similarity differently.

Table 4

Certainty Parameters for Expressions

α Conditional likelihood that the right-hand side of a dependency or
   implication has a particular value (referent) given that the left-hand
   side has a particular value. Applies to dependencies and implications.

β Conditional likelihood that the left-hand side of a dependency or
   implication has a particular value given that the right-hand side
   has a particular value. Applies to dependencies and implications.

γ Degree of certainty or belief that an expression is true. Applies
   to any expression.

τ Degree of typicality of a subset within a set (e.g., robin is a
   typical bird and ostrich is an an atypical bird). Applies to GEN
   and SPEC statements.

σ Degree of similarity of one set to another set. Applies to SIM
   and DIS statements.

ω Frequency of the referent in the domain of the descriptor (e.g., a
   large percentage of birds fly). Applies to any non-relational statement.

δ Dominance of a subset in a set (e.g., chickens are not large
   percentage of birds. but are a large percentage of dominant
   among barnyard fowl). Applies to GEN and SPEC statements.

μ_r Multiplicity of the referent (e.g. many minerals are produced
   by a country like Venezuela). Applies to any non-relational statement.

μ_a Multiplicity of the argument (e.g. many countries
   produce a mineral like oil). Applies to any non-relational statement.

context CX. For example, "chicken" might be compared to "bird" with respect to size or with respect to all its physical features to determine its typicality. For a continuous variable like size, typicality or similarity is determined by computing how close (normalized between 0 and 1) the two values are in the distribution of sizes for the class specified by the context CX (e.g. birds). For discrete variables like "ability to fly", the two concepts either match or not (assigned either 1 or 0), though many discrete variables might better be treated as continuous for comparison processes (e.g., reflecting the degree of flying ability). Typicality or similarity are based on the average score for all the features compared, weighted for their criteriality or importance (Carbonell & Collins, 1973; Collins & Quillian, 1972). We assume the combining function used is something like that proposed by Tversky (1977) where matching features increase the similarity or typicality and mismatching features decrease similarity or typicality.

Frequency (ϕ) and dominance (δ) reflect different ratios that affect the certainty of plausible inferences in systematic ways. Frequency reflects the proportion of members of the argument set that can be characterized by the referent specified. It reflects what "Some" or "All" reflect in logic (e.g., "Some men have arms"), but as a continuous variable between 0 and 1. For the statement "means-of-locomotion (birds)={flying...}," ϕ is the proportion of birds that fly to the total of all birds. The dominance (δ) of a subset within a set applies only to generalization and specialization statements. It reflects the proportion of members of the set that are members of the subset specified in the statement. For example, chickens constitute a high proportion of barnyard fowl, but not of birds in general.

The multiplicity of the referent (μ_r) and multiplicity of the argument (μ_a) are closely related parameters. Suppose somebody thinks that Mineral(Venezuela)={oil...}. The multiplicity of the referent in this case reflects the relative number of minerals (the superordinate of oil) the person thinks Venezuela might have, and the multiplicity of the argument reflects the number of countries (the superordinate of Venezuela) that might produce oil. In this case most people think of both μ_r and μ_a as multiple: i.e., Venezuela produces more than one mineral, and there are other countries that produce oil. In general people don't know about the multiplicity of particular cases, so they derive the multiplicities by inference from more general knowledge: e.g. that countries typically produce more than one mineral, and that any mineral is found in

more than one country. Many descriptors and arguments are thought of as single-valued or low in multiplicity. For example most people think of mines as producing only one mineral (i.e. μ_r=low) though each mineral comes from multiple mines (μ_a=high). For capital (Spain)=Madrid both μ_r is low (Spain has only one capital) and μ_a is low (only one country has Madrid as capital). If the multiplicity of the referent is low, then this corresponds to the fact that the referent is single-valued rather than set-valued. The theory treats the single-valued vs. set-valued distinction as a continuous variable, thereby allowing the degree of certainty derived from inferences to vary continuously with different degrees of multiplicity.

In conclusion, the different primitives in the system can be classified into four groups. The first group are statements representing people's beliefs about the world (e.g. means-of-locomotion(birds)={flying...}). The second group are statements involving relations (i.e. GEN, SPEC, SIM, DIS) representing different relationships between concepts in hierarchies (e.g. canary SPEC bird). The third group are relational expressions called mutual implications and mutual dependencies, that represent people's approximate knowledge about what depends on what, which can be specified with more or less precision. The fourth group are the certainty parameters that act to condition these three kinds of expressions, and which affect the certainty of the different inferences described in the next two sections.

4. STATEMENT TRANSFORMS

The simplest class of inferences in the core theory are called statement transforms. If a person believes some statement, such as that the flowers growing in England[6] include daffodils and roses [i.e., flower-type(England)={daffodils, roses...}], there are eight statement transforms that allow plausible conclusions to be drawn. These eight transforms can be thought of as perturbations of the statement either with respect to the argument hierarchy (starting from England) or the referent hierarchy (starting from daffodils and roses). The transforms of arguments move up (using GEN), down (using SPEC), or sideways (using SIM or DIS) in the argument hierarchy. Similarly the transforms of referents move up, down, or sideways in the referent hierarchy. Thus each of these transforms is a perturbation in one of the two hierarchies.

Let us illustrate the eight transforms of statements in terms of hierarchies for England and roses. Figure 4 shows a part hierarchy for England and a type hierarchy for roses and daffodils that someone might have. If a given person believes that "flower-type(England)={daffodils, roses...}," then Table 5 shows eight conclusions that the person might plausibly draw (assuming other information does not override any of the conclusions).

Insert Figure 4 and Table 5 here

The first GEN inference is that Europe as a whole grows daffodils and roses, which is a kind of induction. This may not be true: Daffodils and roses may be a peculiarity of England, but it is at least plausible that daffodils and roses are widespread throughout Europe. Similarly, for the SPEC operator it is a plausible inference that the county of Surrey in southern England grows roses and daffodils. There is an implicit context (CX) in GEN and SPEC transforms, that will be discussed later.

---

[6]This can be taken to mean flowers that grow outdoors in England which we have simplified to flower-type(England).
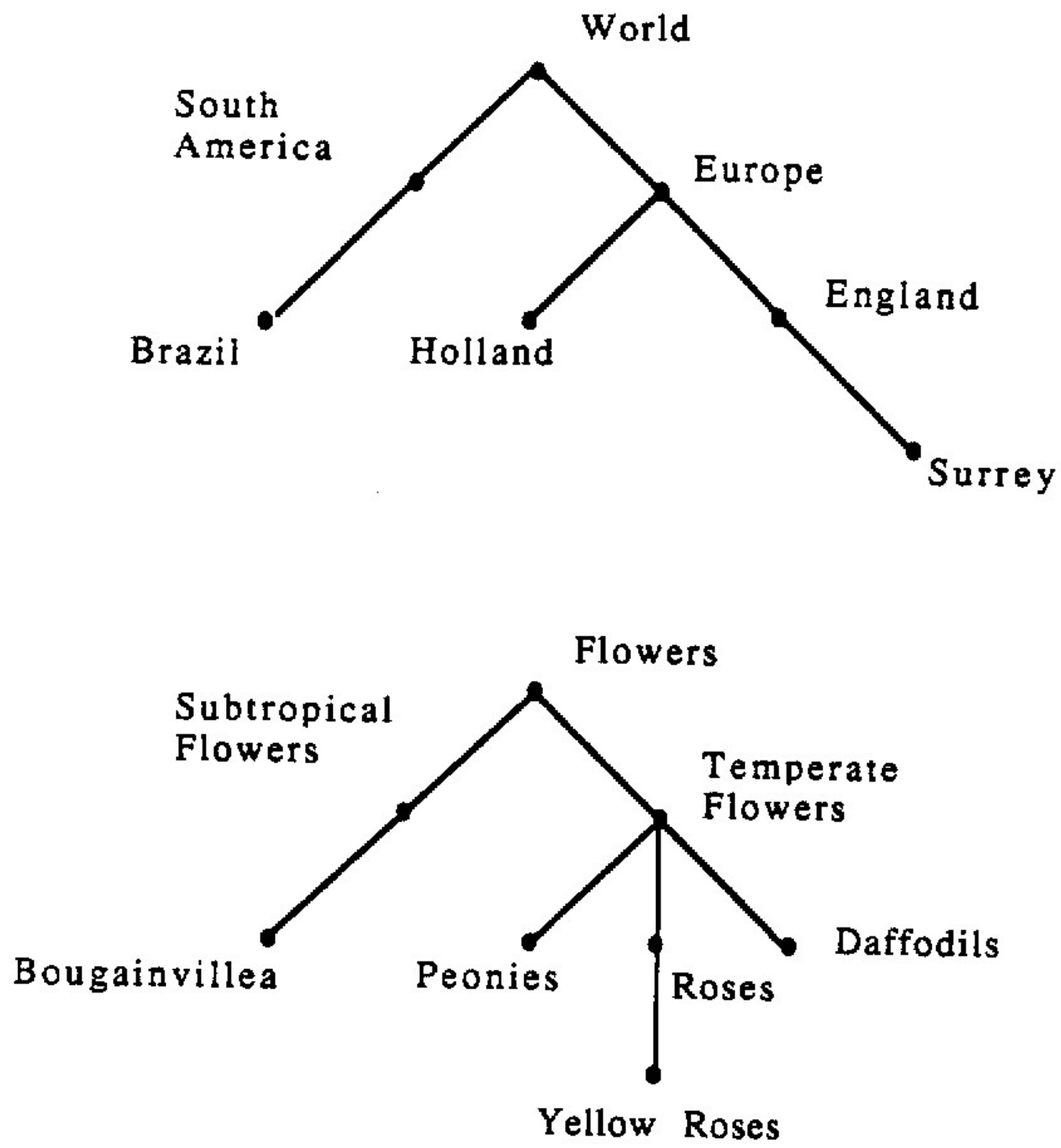
25

Figure 4. Part Hierarchy for England and Type Hierarchy for Roses.

The SIM and DIS inferences are also made in some context. In the case of the transforms of arguments the context might be "countries of the world with respect to the variable climate." Holland is quite similar to England with respect to climate, while Brazil is quite dissimilar. The variables over which the comparison is made may be few or many but people will make the comparison with respect to those variables that they think are most relevant to the question (e.g., whether they grow daffodils in Holland). That is, they base their inference on whatever mutual dependency most constrains the descriptor in question. In this case the flowers grown in a place depend highly on the climate of the place, but hardly at all on the language of the place. Therefore climate is a reasonable variable on which to make the comparison. We will refer to this issue later when we talk about how different parameters affect the certainty of any statement transform.

The transforms of referents are perhaps easiest to understand if you substitute a fictional place like Ruritania for England, because other inferences are not invoked so easily. If one believes they grow daffodils and roses in Ruritania, then one might infer they grow temperate flowers in general there, and yellow roses in particular. It is also reasonable that they grow peonies there, since they are similar to roses and daffodils as to the climates they grow in. But bougainvillea grows in more tropical climates, so it is unlikely to grow in Ruritania (Ruritania is, after all, a small little kingdom and unlikely to encompass different climates——see discussion under multiplicity below). These examples should give a feel for how the transforms of statements are made.

An argument that might be made against the generality of these patterns of inference is that people would draw all sorts of absurd conclusions if they followed these patterns in most cases. For example, since most people know that in general birds fly (i.e., means-of-locomotion(birds)={flying...}), they therefore might conclude by a generalization transform that animals or living things in general fly. Or by a specialization transform that penguins and ostriches fly. Or by a similarity transform that fish or chipmunks fly, if they think fish or chipmunks are in some ways similar to birds. Or by a dissimilarity transform, that insects do not fly because they are so different from birds. Examples like these can be invented indefinitely.

The claim of the theory is that to the degree people do <u>not</u> have contrary

Table 5

Eight Transforms on the Statement
"flower–type(England)={daffodils, roses...}"

Argument Transforms

(1)   GEN   flower–type(Europe)={daffodils, roses...}

(2)   SPEC   flower–type(Surrey)={daffodils, roses...}

(3)   SIM   flower–type(Holland)={daffodils, roses...}

(4)   DIS   flower–type(Brazil)≠{daffodils, roses...}

Referent Transforms

(5)   GEN   flower–type(England)={temperate flowers...}

(6)   SPEC   flower–type(England)={yellow roses...}

(7)   SIM   flower–type(England)={peonies...}

(8)   DIS   flower–type(England)≠{bougainvillea...}

information or make countervailing inferences to override such conclusions, that they indeed will tend to make such inferences. Consider the plight of a young child who has never seen any animals until he is three. If he first meets up with birds and is told that there are many different kinds of animals other than birds, he might well infer that they all fly around in the air like birds. And if he is told that there are different kinds of birds, like bobolinks and starlings, he will infer that they probably fly unless given some reason to think they do not. And if he is shown a little chipmunk that looks a lot like a bird he has seen, he may think it can fly, unless he notices it doesn't have wings. And until he sees insects flying, he might well infer that they do not fly, since they look so different from birds. That is to say the thing that prevents people from making a lot of absurd inferences in our view is the overwhelming dominance of their knowledge about the world: when put into the situation of reasoning about aspects of the world for which they have little knowledge, these kinds of incorrect conclusions are commonplace. That is not to criticize such plausible reasoning: more often than not it leads to correct conclusions, particularly when one has enough information to go on. In order to help the reader see the plausibility of the patterns in the theory, we have tried to choose examples where most readers will not have a lot of relevant information to override the plausibility of the inferences shown.

4.1 Certainty Parameters Affecting Statement Transforms

In this section we will discuss how different certainty parameters affect the various transforms shown in Table 5.

Typicality. Typicality (τ) affects the certainty of any GEN or SPEC transform as shown in Table 6. In transforms of arguments the more typical the subset is of the set in the argument hierarchy, the more certain the inference. For example, in Table 5 inference (1) is more certain the more typical England is as part of Europe, and inference (2) is more certain the more typical Surrey is as part of England.

Insert Table 6 here

In making plausible inferences people compute typicality with respect to those variables, such as climate, that they think flower growing depends on. Thus, if Surrey is thought to have a typical climate for England, and climate is thought to predict the types of flowers grown in a place, then the inference is more certain.

This example reveals the mutual dependency implicit in any statement transform, that has forced us to include a third premise in the statement transforms. The mutual dependency relates the set of variables on which the typicality or similarity judgment is made (e.g., climate or all variables) to the descriptor in question (e.g., flower-type). If the typicality judgment is made considering all variables (as when we said Surrey is a typical English county), the transform will be inherently less certain because of the weak dependency between most variables and any descriptor such as flower-type. Therefore, if you know that Surrey is typical of England in general, it leads to a less certain inference than if you know Surrey is typical of England with respect to climate.

In transforms of referents typicality works the same way, except that it is computed with respect to the subset and its superset in the referent hierarchy. In inference (5) in Table 5, the greater the typicality of daffodils and roses as temperate flowers, the more certain the inference. Similarly in the inference (6), the greater the typicality of yellow roses as roses, the more certain the inference. Pink roses are more typical than yellow roses, and so they are even more likely to be found in England (or Ruritania for that matter). Again the inference is more certain if typicality is measured with respect to the climate in which the flowers are grown.

Similarity Degree of similarity (σ) affects the certainty of any SIM or DIS inference as shown in Table 6.[7] Like typicality, similarity can be computed over all

---

[7]Rips (1975) found that the typicality of a bird affected the certainty of the inference that another type of bird on an island would have the same disease as the first type of bird: in our terms this is a similarity transform. While Rips found that similarity between the two types of birds also affected certainty, there is nothing in our theory that says that the typicality of the bird with the disease should matter. Our view is that this implies there are two chains of supporting inference: one based on the similarity transform, and one based on a generalization (robins have disease x ==> birds in general have disease x) and a corresponding specialization (birds have disease x ==> starlings have disease x). Rips' finding that the typicality of the first bird (robins) mattered more than typicality of the second bird (starlings) may reflect the fact that the generalization is inherently a much less certain inference than the specialization, and so is more affected by the certainty parameters.

Table 6

Effects of Different Parameters on Statement Transforms

Transforms in Table 5		Parameters							Target Node
		τ	σ	α	$\diamond$	$\grave{c}$	μ_r	μ_a	
Argument Transforms	1 GEN	+	0	+	+	+	0	+	Europe
	2 SPEC	+	0	+	+	+	0	0	Surrey
	3 SIM	0	+	+	+	0	0	+	Holland
	4 DIS	0	−	+	−	0	0	−	Brazil
Referent Transforms	5 GEN	+	0	+	+	+	+	0	Temperate Flowers
	6 SPEC	+	0	+	+	+	0	0	Yellow Roses
	7 SIM	0	+	+	+	0	+	0	Peonies
	8 DIS	0	−	+	−	0	−	0	Bougainvillea

Note: As the value of the parameter increases, a + means it has a positive effect on the certainty of the inference and a − means it has a negative effect on the certainty of the inference.

variables or over a subset of variables (e.g., climate) that are particularly relevant in the given context. Degree of similarity increases the certainty of SIM inferences and decreases the certainty of DIS inferences, as would be expected. In Table 5, therefore the inference (3) that Holland has daffodils and roses is more certain the more similar Holland is to England with respect to climate or whatever variables one thinks flowers are related to. The inference (4) that Brazil does not have roses and daffodils is more certain the less similar Brazil is to England. The inference (7) that England has peonies is more certain, the greater the similarity of peonies to both daffodils and roses. The inference (8) that England does not have bougainvillea is more certain, the less similar bougainvillea is to daffodils and roses. More particularly bougainvillea is dissimilar in that it tends to grow in warmer climates than daffodils and roses.

Conditional likelihood. Every statement transform involves an implicit mutual dependency. The inference is always more certain the greater the conditional likelihood (α) between the variables on which typicality or similarity are measured and the variable in question as shown in Table 6. If climate were the variable used for measuring typicality and similarity, the transforms on arguments would be more certain the more the climate of a place constrains the flowers grown in the place. The mutual dependency is slightly different for transforms on referents. They are more certain, the more the climate where flowers grow constrains the places where flowers grow.

Frequency. The frequency (ϕ) of the referent set within the domain of the argument (which is the same as the all, most, or some variable in logic) affects the certainty of all eight inferences, as shown in Table 6. For a particular instance like England, frequency with respect to the argument set only makes sense if you think of England as a set of small parts (about the size of Surrey or Holland) and count the frequency of parts that have daffodils and roses vs. those that do not. The more frequent daffodils and roses are in the parts of England, then all but the DIS inferences are more certain. For example, roses and daffodils are more likely to occur in Holland or Surrey if they are very frequent in England. The two DIS inferences go in the opposite direction. For example, the less frequent are daffodils and roses in England, the more likely bougainvillea will be found there (though this is a very weak inference).

<u>Dominance</u>. Dominance (δ) affects GEN and SPEC inferences as is shown in Table 6. In all cases, the greater the dominance of the subset, the more certain the inference. For example, for (2) if Surrey comprised most of England it would be a more certain inference that it has daffodils and roses, than if it is a very small area in England. Similarly for (6) if yellow roses were the most dominant kind of roses, they would be more likely found in England than if they are a rare type of rose.

<u>Multiplicity</u>. The multiplicity of the referent (μ_r) and the multiplicity of the argument (μ_a) directly affect the eight transforms as shown in Table 6.[8] If England produces many different flower types (μ_r = high), then it makes it more likely that temperate flowers in general grow in England and that peonies, in particular, grow in England (one might even argue it makes yellow roses more likely). However, the negative inference that bougainvillea do not grow in England is less certain if England produces lots of flowers. Similarly, if many countries have daffodils and roses (μ_a is high), it is more likely that Europe in general and Holland in particular have them, and it is less likely that Brazil does not have them.

These two types of multiplicity often determine whether a SIM or DIS inference is invoked. In particular, if a referent is viewed as single valued (e.g. Capital(place) and Weather(place)), then DIS inferences are more certain than SIM inferences. For example, if Capital(Spain)=Madrid and Lisbon DIS Madrid, then probably Capital(Spain)$\neq$Lisbon; or if Weather(Boston)=rain and Sunshine DIS rain, then probably Weather(Boston)$\neq$sunshine. However, to the degree a referent is set-valued (e.g. Minerals(place) and Means-of-locomotion(animal)), then SIM inferences are more certain than DIS inferences. For example, if Mineral(Chile)=copper and Zinc SIM copper, then perhaps Mineral(Chile)=zinc; or if Means-of-locomotion(gazelle)=running and Walking SIM running, then probably Means-of-locomotion(gazelle)=walking.

These examples bring up two caveats. First it is important to understand that

_____

[8]There are also indirect effects of low multiplicity on SIM and DIS transforms where 0's appear in Table 6. If only one or few kinds of flowers were grown in a country (μ_r = low), then it would be less likely that Holland has daffodils and roses (especially if one knows about Holland's tulips), and more likely that Brazil does not have daffodils and roses. Similarly, if very few countries had a particular flower (μ_a = low), then it is less likely that England would have peonies, and more likely that it would not have bougainvillea.

single-valued vs. set-valued referents and arguments are often not clearly distinguished in people's minds, as linguists or logicians might prefer. Instead, there is a continuum between single-valued and set-valued: just because it is raining in a spot doesn't necessarily mean it isn't sunny as well, and just because the capital of Brazil is Brasilia doesn't mean the capital of Brazil isn't Rio de Janeiro as well (where the Congress meets). In a parallel manner, people may get the impression from their school learning that many South American countries produce one mineral product and that is all (i.e. Chile produces copper, Venezuela oil, and Bolivia tin), in which case they may reject other minerals as coming from these countries. So people treat variables as having more or fewer values depending on their understanding of the world.

The other caveat has to do with the way that SIM and DIS inferences are always competing. The competition between SIM and DIS inferences showed up in the goose quacking protocol, where the SIM inference (i.e. ducks quack, and geese are similar to ducks, so maybe geese quack) competed with the DIS inference (ducks quack and geese are different from ducks in their vocal cords, so probably geese do not quack). But often only one of the two (SIM or DIS) is actually invoked, and so whether a SIM or DIS inference is invoked can depend on other variables. If one thinks of the Minerals(Mines) as being single-valued, then one will likely invoke a DIS inference rather than the SIM inference cited above for copper and zinc (e.g. if Mineral(Anaconda mine)=copper and Zinc DIS copper, then Mineral(Anaconda mine)≠zinc). But which inference wins out really does seem to hang on a knife edge. If a person is aware that some mines produce more than one mineral, then they may conclude that the Anaconda mine might very possibly produce zinc as well as copper. In summary both SIM and DIS transforms are appropriate in many cases, and whether a person invokes one or the other often depends on the peculiarities of their knowledge, particularly their knowledge about the multiplicity of the referents and arguments.

4.2 Formal Representation of Statement Transforms[9]

Table 7 shows the formal representations we have developed for each of the eight statement transforms in terms of the variable-valued logic notation of Michalski (1983). Most of the examples shown are from protocols we have collected (Collins, 1978b), some of which appear in the first section of this paper. We will briefly describe each of the examples.

Insert Table 7 here

We can illustrate GEN-based argument transforms with the inference that if chickens have gizzards, then birds in general may have gizzards. The first premise, represents the belief that chickens have gizzards. presumably almost all chickens have gizzards so the frequency (ϕ) and the certainty (γ) are high. Presumably μ_a is also high because any internal organ tends to occur in many different animals. The second premise represents the belief that chickens are birds, and that they are typical with respect to their biological characteristics. As we pointed out earlier, the dominance (δ) of chickens among birds is low. The third premise states that the internal organs of a bird depend highly on the biological characteristics of the bird. The conclusion that birds in general have gizzards is fairly certain given the high values of the critical variables.

SPEC-based argument transforms are illustrated by an example from the beginning of the paper where the respondent inferred that the Andes might be in Uruguay. The respondent believed that the Andes are in most South American countries, so frequency (ϕ) was moderately high. With respect to the second premise, Uruguay is a typical South American country, which increases the likelihood that the Andes would be found there. But its low dominance (δ) in terms of the proportion of South America that Uruguay comprises makes the inference less likely. With respect to the third premise, the fact that Uruguay is typical of South American countries in general only weakly predicts that it will include the Andes mountains. Altogether, the

---

[9]This section can safely be skipped by readers.

Table 7

Formal Representations of Statement Transforms*

(1) GEN-based argument transforms

$$d(a)=r: \quad \gamma, \phi, \mu_a$$
$$a' \text{ GEN } a \text{ in CX } (a';D(a')): \quad \tau, \gamma_2, \delta$$
$$D(a') \quad <----> \quad d(a'): \alpha, \gamma_3$$

$$d(a')=r: \quad \gamma = f(\gamma_1, \phi, \mu_a, \tau, \gamma_2, \delta, \alpha, \gamma_3)$$

Internal organ (chicken) = {gizzard ...}: γ_1=high, ϕ=high, μ_a=indeterminate
Bird GEN chicken in CX (bird; biological characteristics(bird)):
  τ=high, γ_2=high, δ=low
Biological characteristics (bird) <----> Internal organ (bird):
   α=high, γ_3=high

Internal organ (bird) = {gizzard ...}: γ=high

(2) SPEC-based argument transforms

$$d(a)=r: \quad \gamma_1, \phi$$
$$a' \text{ SPEC } a \text{ in CX } (a, D(a)): \quad \tau, \gamma_2, \delta$$
$$D(a) <----> d(a): \quad \alpha, \gamma_3$$

$$d(a')=r: \quad \gamma = f(\gamma_1, \phi, \tau, \gamma_2, \delta, \alpha, \gamma_3)$$

Mountains (S.A. country) = {Andes ...}: γ_1=high, ϕ=high, μ_a=indeterminate
Uruguay SPEC S.A. country in CX (S.A. country; characteristics(S.A. country)):
  τ=high, γ_2=high, δ=low
Characteristics (S.A. country) <----> Mountains (S.A. country):
  α=moderate, γ_3=high

Mountains (Uruguay) = {Andes ...}: γ=moderate

* D and A represent superordinates of d and a respectively.

(3) SIM—based argument transforms

$d(a) = r. \gamma_1, \phi, \mu_a$
a' SIM a in CX (A: D(A)). $\sigma. \gamma_2$
$D(A) <---> d(A). \alpha, \gamma_3$
a, a' SPEC A: γ_4, γ_5
_____

$d(a') = r: \gamma = f(\gamma_1, \phi, \mu_a, \sigma, \gamma_2, \alpha, \gamma_3, \gamma_4, \gamma_5)$

Livestock (West Texas) = {cattle ...}: γ_1=high, ϕ=high, μ_a = high
Chaco SIM West Texas in CX (region; vegetation(region)):
    σ=moderate, γ_2=moderate
Vegetation (region) <----> Livestock (region): α=high, γ_3=high
West Texas, Chaco SPEC region. γ_4=high, γ_5=high
_____

Livestock (Chaco) = {cattle ...}. γ=moderate

(4) DIS—based argument transforms

$d(a) = r: \gamma_1, \phi, \mu_a$
a' DIS a in CX(A; D(A)). σ, γ_2
$D(A) <----> d(A): \alpha, \gamma_2$
a, a' SPEC A: γ_4, γ_5
_____

$d(a') \neq r. \gamma = f(\gamma_1, \phi, \mu_a, \sigma, \gamma_2, \alpha, \gamma_3, \gamma_4, \gamma_5)$

Sound (duck) = quack: γ_1=high, ϕ=high, μ_a=low
Goose DIS duck in CX (bird; vocal cords (bird)):
    σ=low, γ_2=moderate
Vocal cords (bird) <----> Sound (bird). α=high, γ_3=low
Duck, goose SPEC bird: γ_4=high, γ_5=high
_____

Sound (goose) $\neq$ quack: γ=low

(5) GEN-based referent transforms

$d(a) = \{r \ldots\}. \quad \gamma_1, \phi, \mu_r$
r' GEN r in CX(d; D(d)): τ, γ_2, δ
D(d) <----> A(d). α, γ_3
a SPEC A. γ_4
- -
$d(a) = \{r' \ldots\}. \quad \gamma = f(\gamma_1, \phi, \mu_r, \tau, \gamma_2, \alpha, \gamma_3, \gamma_4)$

Agricultural product (Honduras) = {bananas …}:
    γ_1=unknown, ϕ=high, μ_r=high
Tropical fruits GEN bananas in CX (agricultural products,
    climate(agricultural products)): τ=high, γ_2=high, δ=low
Climate (agricultural products) <----> Place (agricultural products):
    α=high, γ_3=high
Honduras SPEC place: γ_4=high
- -
Agricultural products (Honduras)={tropical fruits…}: γ =moderate

(6) SPEC-based referent transforms

$d(a) = \{r \ldots\}: \quad \gamma_1, \phi$
r' SPEC r in CX(d; D(d)). τ, γ_2, δ
D(d) <----> A(d): α, γ_3
a SPEC A: γ_4
- -
$d(a) = \{r' \ldots\}. \quad \gamma = f(\gamma_1, \phi, \tau, \gamma_2, \delta, \alpha, \gamma_3, \gamma_4)$

Minerals (South Africa) = {diamonds…}. γ_1=high, ϕ=high
Industrial diamonds SPEC diamonds in CX (minerals, characteristics(minerals)):
    τ =high, γ_2 =high, δ =high
Characteristics(minerals) <----> Place (minerals).
    α =moderate. γ_3 =high
South Africa SPEC place. γ_4=high
- -
Minerals (South Africa) = {industrial diamonds …}. γ =high

(7) SIM-based referent transforms

$$d(a) = \{r...\}: \quad \gamma_1, \, \Phi, \, \mu_r$$
$$r' \text{ SIM } r \text{ in } CX(d; D(d)): \quad \sigma, \, \gamma_2$$
$$D(d) <----> A(d): \quad \alpha, \, \gamma_3$$
$$a \text{ SPEC } A: \quad \gamma_4$$

$$d(a) = \{r'...\}: \quad \gamma = f(\gamma_1, \, \Phi, \, \mu_r, \, \sigma, \, \gamma_2, \, \alpha, \, \gamma_3, \, \gamma_4)$$

Sound (wolf) = {howl...}: γ_1=high, Φ=high, μ_r=low
Bark SIM howl in CX(sound; means of production(sound)):
    σ =high, γ_2=high
Means of production (sound) <----> animal (sound): α=high, γ_3=high
Wolf SPEC animal: γ_4=high

Sound (wolf) = {bark...}: γ=moderate

(8) DIS-based referent transforms

$$d(a) = \{r...\}: \quad \gamma_1, \, \Phi, \, \mu_r$$
$$r' \text{ DIS } r \text{ in } CX(d; D(d)): \quad \sigma, \, \gamma_2$$
$$D(d) <----> A(d): \quad \alpha, \, \gamma_3$$
$$a \text{ SPEC } A: \quad \gamma_4$$

$$d(a) \neq \{r'...\}: \quad \gamma = f(\gamma_1, \, \Phi, \, \mu_r, \, \sigma, \, \gamma_2, \, \alpha, \, \gamma_3, \, \gamma_4)$$

Color (Princess phones) = {white, pink, yellow...}: γ_1=high, Φ=high, μ_r = moderate
Black DIS white & pink & yellow in CX (color; lightness(color)):
    σ=low, γ_2=high
Lightness (color) <----> phone type (color): α=low, γ_3=high
Princess phone SPEC phone: γ_4=high

Color (Princess phones) $\neq$ {black...}: γ=moderate

inference is fairly uncertain given the moderate frequency and the low dominance of Uruguay.

We can illustrate SIM-based argument transforms with the Chaco protocol from the beginning of the paper, where the respondent inferred that the Chaco might produce cattle given that west Texas did. In the first premise, the frequency (ϕ) with which different parts of west Texas have cattle is high, and the multiplicity (μ_σ) of places with cattle is high, both of which make the inference more likely. The second premise asserts that the Chaco is at least moderately similar to west Texas in vegetation (or whatever variables the respondent had in mind). The third premise relates vegetation of a region to its livestock, which is a strong relation, given that cattle will usually be raised where the vegetation will support them. The fourth premise merely establishes the fact that west Texas and the Chaco are regions, in support of the second and third premises. The conclusion is only moderate in certainty, given our assumption of uncertainty about how similar the Chaco and west Texas are.

To illustrate DIS-based argument transforms, we chose the example from the protocol shown earlier as to whether a goose quacks. The first premise reflects the respondent's belief that ducks quack, which was very certain. Though almost all ducks quacks (ϕ is high), very few other animals quack (μ_σ is low), which makes the DIS inference more certain. The second premise states the belief that ducks and geese are dissimilar in their vocal cords which the respondent must have been at least a bit uncertain about (hence the low certainty assigned to the statement). The third premise relates the sound a bird makes to its vocal cords, which also must have been an uncertain belief given that it is not true. The certainty of the conclusion that geese do not quack should have been fairly low (though another inference led to the same conclusion in the actual protocol).

We have created an example to illustrate GEN-based referent transforms. The first premise asserts that Honduras produces bananas among other things (the multiplicity (μ_r) of agricultural products is high). Bananas are a fairly typical tropical fruit in terms of the climates where they are grown, as the second premise states. The third premise asserts that the climate appropriate for agricultural products constrains the places where they are grown fairly strongly. The conclusion follows

with moderate certainty that Honduras produces many tropical fruits, such as mangos and coconuts.

We also created the example of SPEC-based referent transforms. The first premise states that South Africa produces diamonds. Industrial diamonds are a kind of low quality diamond (used in drills) and they must be fairly dominant (δ) among diamonds given their low quality, though they are not particularly typical of what we think of as diamonds. Here is a case where high dominance compensates for low typicality. The third premise is somewhat irrelevant since the typicality is low. But the inference that South Africa produces industrial diamonds is quite certain given the high dominance of industrial diamonds among diamonds.

The example of SIM-based referent transforms is drawn from a protocol where the respondent, when asked whether wolves could bark, inferred they probably could (Collins, 1978b). One of his inferences derived from the fact that he knew wolves could howl, with both high frequency and certainty (but low multiplicity (μ_r) because most animals only make one or two sounds). He also thought that barking was similar to howling in terms of the way the sound is produced (a howl, as it were, is a sustained bark). Furthermore, the means of production of a sound constrains the type of animals that can make that sound, as the third premise states. It follows then with at least moderate certainty that a wolf can bark.

The example of DIS-based referent transforms is from a protocol where the respondent was asked if there are black princess telephones (Collins, 1978b). The respondent could remember seeing white, pink and yellow princess phones, as the first premise states. Here the frequency (ϕ) of these colors among those she had seen seemed quite high, which counts against the possibility of black princess phones. But the multiplicity of different colors among phones (μ_r) is moderate, which counts for the possibility of black princess phones. The second premise reflects the fact that black is quite dissimilar to those colors in terms of lightness. The third premise states that knowing the lightness only somewhat constrains the type of phone (α is low). The conclusion that princess phones are not black is uncertain given the low α in the third premise and the moderate μ_r in the first premise.

5. INFERENCES BASED ON IMPLICATIONS AND DEPENDENCIES

The previous section illustrated the systematic patterns by which one statement can be transformed into another. The pattern of inferences based on mutual implications and dependencies is somewhat more complicated, but is also quite systematic. There are three basic classes of these inferences: (a) derivations from mutual implications and dependencies, where a statement is derived from an implication or dependency, (b) transitivity inferences, where a new implication or dependency is derived from a given pair of implications or dependencies, and (c) argument or referent transforms based on implication or dependency, that parallel the statement transforms shown in the previous section. In this section we give the formal representation for these inference patterns together with an example of each.

5.1 Derivations from Implications and Dependencies

Table 8 illustrates the two types of *derivation from mutual implication* that occurred in the protocols shown at the beginning of the paper. The positive derivation illustrates how multiple conditions were ANDed together (i.e., a warm climate, heavy rainfall, and flat terrain) as predictors of rice growing. The belief that Florida has all three leads to a prediction that rice will be grown there. In the actual protocol the respondent was unsure about rainfall in Florida, and so concluded that rice would be grown if there was enough rain (i.e., Rainfall(Florida) = heavy <===> Product(Florida) = {rice...}).

Insert Table 8 here

The negative derivation illustrates the fact that if any of the variables on one side of a mutual implication that are ANDed together do not have the appropriate values, then you can conclude that the variable on the other side does not have the value assumed in the mutual implication. In the example, because the Llanos did not have reliable rainfall, the respondent concluded that the Llanos probably did not produce coffee. If variables are ORed together (e.g., either heavy rainfall or irrigation

35

Table 8

Formal Representations of Derivations from Mutual Implication

Positive Derivation

$$d_1(a) = r_1 \iff d_2(a) = r_2 : \alpha, \gamma_1$$
$$d_1(a') = r_1 : \phi, \gamma_2$$
$$\underline{a' = \text{SPEC } a : \gamma_3}$$
$$d_2(a') = r_2 : \gamma = f(\alpha, \gamma_1, \phi, \gamma_2, \gamma_3)$$

Climate(place) = warm & Rainfall(place) = heavy & Terrain(place) = flat $\iff$
      Product(place) = {rice...} : α = high, γ_1 = certain
Climate(Florida) = warm : ϕ_1 = moderately high, γ_2 = certain
Rainfall(Florida) = heavy : ϕ_2 = moderate, γ_3 = uncertain
Terrain(Florida) = flat : ϕ_3 = high, γ_4 = certain
Florida SPEC place : γ_5 = certain
Product(Florida) = {rice...} : γ = uncertain

Negative Derivation

$$d_1(a) = r_1 \iff d_2(a) = r_2 : \alpha, \gamma_1$$
$$d_1(a') \neq r_1 : \phi, \gamma_2, \mu_r$$
$$\underline{a' \text{ SPEC } a : \gamma_3}$$
$$d_2(a') \neq r_2 : \gamma = f(\alpha, \gamma_1, \phi, \gamma_2, \mu_r, \gamma_3)$$

Rainfall(place) = reliable & climate(place) = subtropical $\iff$
      Product(place) = {coffee...} : α = moderate, γ_1 = certain
Rainfall(Llanos) $\neq$ reliable : ϕ = high, γ_3 = fairly certain, μ_r = low
Llanos SPEC place : γ_3 = certain
Product(Llanos) $\neq$ {coffee...} : γ = fairly certain

are needed for growing rice) a different pattern holds: having one or the other predicts rice is grown and having neither predicts no rice is grown.

Table 9 shows the equivalent representations for derivations from mutual dependencies. The inference patterns are different for positive and negative dependencies, so we have separated them in the table. It is possible to draw a negative conclusion from a mutual implication simply by negating the second premise and the conclusion in either of the patterns shown.

Insert Table 9 here

The positive dependency represents the case where as one variable increases, the other variable also increases. In the formal analysis we have denoted the entire range of both variables by three values: high, medium, and low. When a positive dependency holds, if the values of the first variable is high, medium, or low, the value of the second variable will also be high, medium, or low, respectively. This is the weakest kind of derivation possible from a mutual dependency: In the example, if a person knows that the temperature of air predicts the water holding capacity of air, and he knows that temperature of the air outside is warm, then he can infer that the air outside could hold a lot of moisture. People make this kind of weak inference very frequently in reasoning about such variables (Collins & Gentner, 1987; Stevens & Collins, 1980).

The pattern for the negative dependency is reversed: if the value of one variable is high, the other is low, and vice versa. We have illustrated the derivation from a negative dependency in terms of a more precise dependency between two variables. If a person believes that the latitude of a place varies negatively (and linearly) with the temperature of the place, and also that the average temperature is near 85 degrees at the equator and 0 degrees at the poles, then he might conclude that a place like Lima, Peru, that is about 10 degrees from the equator, has an average temperature of about 75 degrees. People have both more and less precise notions of how variables interact, and we have tried to preserve flexibility within our representation for handling these different degrees of precision.

Table 9

Formal Representations of Derivations from Mutual Dependencies

Derivation from Positive Dependency

$d_1(a) <--^+--> d_2(a) : \alpha, \gamma_1$
$d_1(a') =$ high, medium, low : ϕ, γ_2
<u>a' SPEC $a : \gamma_3$</u>
$d_2(a') =$ high, medium, low : $\gamma = f(\alpha, \gamma_1, \phi, \gamma_2, \gamma_3)$

Temperature(air) $<--^+-->$ Water holding capacity(air) : $\alpha =$ high, $\gamma_1 =$ certain
Temperature(air outside) = high : $\phi =$ high, $\gamma_2 =$ certain
<u>Air outside SPEC air : $\gamma_3 =$ certain</u>
Water holding capacity(air outside) = high : $\gamma =$ certain

Derivation from Negative Dependency

$d_1(a) <--^---> d_2(a) : \alpha, \gamma_1$
$d_1(a') =$ high, medium, low : ϕ, γ_2
<u>a' SPEC $a : \gamma_3$</u>
$d_2(a') =$ low, medium, high : $\gamma = f(\alpha, \gamma_1, \phi, \gamma_2, \gamma_3)$

Abs. Val. Latitude(place) $<--^---->$ Aver. Temperature(place): linear;
    $0°, 85°, 90°, 0°, \alpha =$ moderate. $\gamma_1 =$ certain

Abs. Val. Latitude(Lima Peru) = $10°$: $\phi =$ high. $\gamma_2 =$ fairly certain
<u>Lima Peru SPEC place : $\gamma_3 =$ certain</u>

Aver. Temperature(Lima Peru $= 75°$. $\gamma =$ moderately certain

5.2 Transitivity Inferences

Table 10 shows two forms of a transitive inference, one based on mutual implication and the other based on mutual dependency. The example for mutual implication states that if a person believes an average temperature of 85 degrees implies a place is equatorial, and that if a place is equatorial it will tend to have high humidity, then he can infer that if the average temperature of a place is 85 degrees it will tend to have high humidity, and vice versa. This example illustrates the way people often confuse causality and diagnosticity in their understanding.[10] If one were to write the causal links for this example, it would probably go from equatorial latitude to high temperature to high humidity. But people do not systematically make a distinction between causal and diagnostic links, nor do they store things in such a systematic order. For example, they may know that equatorial places, such as jungles, have high humidity and not link it explicitly to their high temperature. Thus, the inference in this example derives a more direct link (temperature <==> humidity) from a less direct link (latitude <==> humidity). It also should be noted that the diagnostic link in the first implication (temperature => latitude) may be more constraining than the causal link (latitude => temperature). That is, there are probably more equatorial places where the average temperature is not 85 degrees (e.g. Ecuador), than places where the average temperature is 85 degrees but are not equatorial.

Insert Table 10 here

The example for a transitivity inference on mutual dependency illustrates how people reason about economics (Salter, 1983). Salter asked subjects questions, such as what is the effect of an increase in interest rates on the inflation rate of a country. People gave him chains of inferences like the one shown. if interest rates increase, then growth in the money supply will decrease, and that in turn will cause the inflation rate to decrease (the latter is a positive dependency). So an increase in

---

[10]This is not to say that where people do make a clear distinction between causality and diagnosticity, as in the examples cited by Tversky and Kahneman (1980), that they do not treat α and β asymmetrically, giving preference to causal links.

Table 10

Formal Representations of Transitivity Inferences

On Mutual Implication

$d_1(a) = r_1 <==> d_2(a) = r_2$: α_1, β_1, γ_1

$d_2(a) = r_2 <==> d_3(a) = r_3$. α_2, β_2, γ_2

$d_1(a) = r_1 <==> d_3(a) = r_3$: $\alpha = f(\alpha_1, \alpha_2)$, $\beta = f(\beta_1, \beta_2)$ $\gamma = f(\gamma_1, \gamma_2)$

Aver. Temperature(place) = 85° <==> Latitude(place) = equatorial :
    α_1 = high, β_1 = fairly high, γ_1 = certain

Latitude(place) = equatorial <==> Abs. humidity(place) = high :
    α_2 = high, β_2 = moderate, γ_2 = certain

Aver. Temperature(place) = 85° <==> Abs. Humidity(place) = high :
    α = high, β = low, γ = certain

On Mutual Dependency

$d_1(a) <--> d_2(a)$: α_1, β_1, γ_1

$d_2(a) <--> d_3(a)$: α_2, β_2, γ_2

$d_1(a) <--> d_3(a)$: $\alpha = f(\alpha_1, \alpha_2)$, $\beta = f(\beta_1, \beta_2)$ $\gamma = f(\gamma_1, \gamma_2)$

Interest rates(country) $<-^{-}->$ Money supply growth(country):
    α_1 = high, β_1 = moderate, γ_1 = certain

Money supply growth(country) $<-^{+}->$ Inflation rate(country):
    α_2 = high, β_2 = high, γ_2 = certain

Interest rates(country) $<-^{-}->$ Inflation rate (country):
    α_3 = high, β_3 = low, γ_3 = certain

interest rates will lead to a decrease in the inflation rate. This kind of reasoning is a major way that people construct new mutual implications and dependencies.

5.3 Transforms based on Implications and Dependencies

Tables 11 and 12 show a set of transforms based on mutual implications that follow the same pattern as the statement transforms in the previous section. Table 11 shows four referent transforms that parallel the last four statement transforms shown in Tables 5 and 7. (In fact there is a quite direct equivalence, because any statement can be transformed into a mutual implication in the following way: Flowers (England) = {daffodils...} goes into type(place) = England <==> flowers(place) = {daffodils...}, or more generally. d(a) =r goes into type(A)=a <==> d(A) = r.) We have represented the three positive transforms (i.e. generalization, specialization, and similarity) in the rule at the top. with the three alternatives shown (GEN, SPEC, and SIM) where they occur in the rule. The typicality parameter (τ) is associated with the GEN and SPEC transforms, and the similarity parameter (σ) with the SIM transform. The example omits the certainty parameters for simplicity. In English the example states the following. given the belief that if a place is subtropical. it is likely to produce oranges, this implies that if a place is subtropical, it is likely to produce citrus fruits (a generalization). or navel oranges (a specialization). or grapefruit (a similarity transform). The dissimilarity transform at the bottom follows the same pattern: if you think that subtropical places produce oranges, and apples are dissimilar to oranges with respect to their growing conditions, then probably subtropical places do not produce apples.

Insert Table 11 here

Table 12 shows the corresponding four types (i.e.. GEN. SPEC, SIM, and DIS) of argument transforms. These correspond to the first four statement transforms shown in Tables 5 and 7. We illustrate the four with a demographic example: if one believes that men who live in the tropics have a short life expectancy and that farmers are typical of men in terms of their demographic characteristics, then one can plausibly infer that farmers have a short life expectancy if they live in the tropics. Similarly

Table 11

Formal Representations of Referent Transforms based on Mutual Implications

Positive Transforms

$$d_1(a) = r_1 <==> d_2(a) = r_2 : \alpha_1, \gamma_1, \mu_r$$

$$r'_2 \begin{Bmatrix} GEN \\ SPEC \\ SIM \end{Bmatrix} r_2 \text{ in } CX(d_2, D(d_2)) : \{\tau/\sigma\}, \gamma_2$$

$$\underline{D(d_2) <--> A(d_2) : \alpha_2, \gamma_3}$$

$$d_1(a) = r_1 <==> d_2(a) = r'_2 : \gamma = f(\alpha_1, \gamma_1, \mu_r, \{\tau/\sigma\}, \gamma_2, \alpha_2, \gamma_3)$$

Climate(place) = subtropical <==> Fruit(place) = {oranges...}

$$\begin{Bmatrix} \text{Citrus fruits} \\ \text{Navel oranges} \\ \text{Grapefruit} \end{Bmatrix} \begin{Bmatrix} GEN \\ SPEC \\ SIM \end{Bmatrix} \text{oranges in CX (fruit; growing conditions(fruit))}$$

Growing conditions(fruit) <--> Place(fruit)

Climate(place) = subtropical <==> Fruit(place) = $\begin{Bmatrix} \{Citrus\ fruit...\} \\ \{Naval\ oranges...\} \\ \{Grapefruit...\} \end{Bmatrix}$

Negative Transform

$$d_1(a) = r_1 <==> d_2(a) = r_2 : \alpha_1, \gamma_1, \mu_r$$

$$r'_2 \text{ DIS } r_2 \text{ in CX } (d_2; D(d_2)) : \sigma, \gamma_2$$

$$\underline{D(d_2) <--> A(d_2) : \alpha_2, \gamma_3}$$

$$d_1(a) = r_1 <==> d_2(a) \neq r'_2 : \gamma = f(\alpha_1, \gamma_1, \mu_r, \sigma, \gamma_2, \alpha_2, \gamma_3)$$

Climate(place) = subtropical <==> Fruit(place) = {oranges...}

Apples DIS oranges in CX (fruit; growing conditions (fruit))

Growing conditions(fruit) <--> Place (fruit)

Climate(place) = subtropical <==> Fruit(place) $\neq$ {apples...}

one can infer that people in general and women (because they are similar to men in their demographic characteristics) have short life expectancy in the tropics. Finally, one might conclude that birds do not have a short life expectancy in the tropics, if one thinks they are dissimilar to men in their demographic characteristics.

Insert Table 12 here

Table 13 shows the corresponding positive transforms based on mutual dependencies. We have illustrated these with another example from economics: if one believes that the business tax rate in a state negatively impacts the amount of investment in the state, then one might generalize this relationship to any governmental unit, or particularize it to Illinois, or conclude that it would also apply to Canadian provinces. There is really no negative transform based on dissimilarity that corresponds to these three positive transforms. For example, if one believes that communist countries are quite dissimilar from states in their economics, the most one can conclude is that there is no negative relation between the business tax rate (if there were one) and the amount of investment; that is to say, no conclusion can be drawn. In such a case we just omit the form from the theory, because the theory does not specify conclusions that cannot be drawn. Similarly, there can be no referent transforms based on mutual dependencies, because they do not involve a referent term.

Insert Table 13 here

Table 12

Formal Representations of Argument Transforms based on Mutual Implications

Positive Transforms

$$d_1(a) = r_1 <==> d_2(a) = r_2 : \alpha_1, \gamma_1$$

$$a' \begin{Bmatrix} GEN \\ SPEC \\ SIM \end{Bmatrix} (a) \text{ in CX } (A; d_3(A)) : \{\tau/\sigma\}, \gamma_2$$

$$\underline{d_3(A) <---> d_2(A) : \alpha_2, \gamma_3}$$

$$d_1(a') = r_1 <===> d_2(a') = r_2 : \gamma = f(\alpha_1, \gamma_1, \{\tau/\sigma\}, \gamma_2, \alpha_2, \gamma_3)$$

Habitat(man) = tropics <==> Life expectancy (man) = short

$$Man \begin{Bmatrix} GEN & farmer \\ SPEC & person \\ SIM & woman \end{Bmatrix} \text{ in CX (people; demographic characteristics(people))}$$

Demographic characteristics(people) <--> Life expectancy(people)

$$Habitat \begin{Bmatrix} (farmer) \\ (person) \\ (woman) \end{Bmatrix} = tropics <==> Life expectancy \begin{Bmatrix} (farmer) \\ (person) \\ (woman) \end{Bmatrix} = short$$

Negative Transforms

$$d_1(a) = r_1 <==> d_2(a) = r_2 : \alpha_1, \gamma_1$$

$$a' \text{ DIS } a \text{ in CX}(A; d_3(A)) : \sigma, \gamma_2$$

$$\underline{d_3(A) <--> d_2(A) : \alpha_2, \gamma_3}$$

$$d_1(a') = r_1 <==> d_2(a') \neq r_2 : f(\alpha_1, \gamma_1, \sigma, \gamma_2, \alpha_2, \gamma_3)$$

Habitat(man) = tropics <==> Life expectancy(man) = short

Man DIS bird in CX (animals, demographic characteristics (animals))

Demographic characteristics(animals) <--> Life expectancy (animals)

Habitat(birds) = tropics<==> Life expectancy(birds) $\neq$ short

Table 13

Formal Representations of Argument Transforms based on Mutual Dependencies

Positive Transforms

$$d_1(a) <--> d_2(a) \quad \alpha_1, \gamma_1$$

$$a' \begin{cases} GEN \\ SPEC \\ SIM \end{cases} a \text{ in } CX \ (A; d_3(A)) \ : \ \{\tau/\sigma\}, \ \gamma_2$$

$$\underline{d_3(A) <--> d_2(A) \quad \alpha_2, \gamma_3}$$

$$d_1(a') <--> d_2(a') \ : \ \gamma = f(\alpha_1, \gamma_1, \{\tau/\sigma\}, \gamma_2, \alpha_2, \gamma_3)$$

Business tax rate (state) $<-^--->$ Amount of investment (state))

$$\begin{cases} \text{Government unit} \\ \text{Illinois} \\ \text{Province} \end{cases} \begin{cases} GEN \\ SPEC \\ SIM \end{cases} \text{state in CX (place; economics (place))}$$

$$\underline{\text{Economics(place)} <-^--> \text{Amount of investment(place)}}$$

Business tax rate $\begin{cases} \text{(government unit)} \\ \text{(Illinois)} \\ \text{(province)} \end{cases}$ $<-^--->$ Amount of investment $\begin{cases} \text{(government unit)} \\ \text{(Illinois)} \\ \text{(province)} \end{cases}$

6. CONCLUSION

We conclude with a few comments about the methodology being used to construct and test the theory. It is difficult for experimental psychologists to find experiments that address the processes that people use to answer everyday questions. The problem is that cognitive psychology's methods are limited for the most part to percent correct and response time measures. Trying to understand the processing in the human mind with these two measures is like trying to conduct a surgical operation with a hammer and chisel. The tools are inappropriate for the questions involved. Cognitive psychologists manage to carry off some clever operations despite their tools, but at the same time they should be looking for finer—grain tools.

The methodology of fitting the arguments made in a set of human responses to a minimal set of argument forms is an attempt to develop one such fine—grain method. The method attempts to balance the constraints necessary to produce consistent structures. It is not a hypothesis—testing method: the forms used to fit the data are for the most part derived from the data. The difficulty of the data analysis is to find the optimal decomposition of the argument forms, so that the set of forms is in some sense minimal (i.e. there are not a large set of forms that share subparts). In other words the difficulty is to extract all the regularities from the data. Suffice to say we have only partially succeeded in this endeavor.

There are real limitations to the methodology, just as there are limitations when an astronomer studies the sky using only the visible spectrum. Some of the problems with protocol analysis as developed by Newell and Simon (1972) apply to the analysis of people's answers that the theory is based on: it is both a post—hoc analysis and a highly inferential analysis. Unlike protocol analysis the method used here does not interfere with normal processing; people just answer in a way that is the normal conversational mode. Some psychologists worry that answering so many questions may force people into a special mode of answering questions, but the patterns of inference appear to be the same in the teaching dialogues we have collected. Even if people are more articulate about their reasoning in this kind of setting, they are not inventing new modes of reasoning for the occasion.

Another difficulty in constructing a theory of plausible reasoning from analyzing

actual cases of human reasoning is that the theory is likely to be underconstrained. That is to say, there may be many cases where people could employ a particular reasoning pattern, but do not because of other constraints on its invocation. As it stands now, the only constraints we place on the invocation of any inference pattern is that its premises be satisfied and that its certainty parameters not drive the conclusion below some threshold level of certainty. But there may well be other factors that constrain the invocation of any inference pattern.

A more serious limitation to the method is its bias against non-verbal processes. We can illustrate this with the following protocol.

Q. How many piano tuners do you think are in New York City?

DK. Well now let me think. How many people are there in New York City anyway? If you think about the whole area, I suppose there may be 10 or 12 million people. You don't need a lot of piano tuners to keep a whole city in tune. Maybe a thousand?

Q. Why do you guess a thousand?

DK. Well, let's think. If there are 12 million people in the city. How many households might there be, and what proportion of them would have pianos and then how many...? We're talking about employed piano tuners I suppose. A piano tuner must need to do a whole lot of pianos just to keep bread on the table. Ah. A thousand is beginning to sound a little high because if one guy does a whole lot of pianos, he'll cover a lot of ground. There must be fewer piano tuners than there are doctors in the city. They can service more pianos, and they are fewer and farther between. It's just a matter of what feels right and might be off by an order of magnitude either way. Maybe 300.

In the protocol the respondent attempts to carry out a means-ends analysis of the problem (Newell & Simon, 1972), but he never carries it through.[11] Rather he seems first to pull the number 1000 out of the air, which he then revises down to 300. The number 1000 could have been derived from any number of non-verbal processes. In any case, whatever process was used, there is no trace in the protocol of it. Such non-verbal processes may be equally systematic as the verbal processes that are so omnipresent; it is just that they do not show up in the protocols. The danger John Seely Brown and Jonathan Baron (personal communication) point out is that the verbal

---

[11]Directed search techniques, like means-ends analysis are beyond the scope of the theory, though they often utilize information obtained from the kind of automatic inferences included in the theory.

protocols may be rationalizations for answers arrived at by some other process. Our intuition is different. It is that the answers frequently follow from both from verbal and non-verbal reasoning processes and that these are weighed together in responding. In answering the piano tuner question some subjects have actually carried through a verbal process (in particular the means-ends analysis the subject quoted above started or a functional analogy) and the answers they derived clearly followed from the verbal process. If our analysis is correct, the responses shown in the five protocols at the beginning of the paper follow at least in part from a verbal process. It is certainly true that non-verbal processes will not be as visible in the responses, though some subjects certainly allude to them (Collins 1978a,b). But their existence does not negate the ubiquity of patterns we have identified in people's reasoning. Our position then is that while there may be additional processes used to answer questions that are not apparent in the responses, nevertheless the processes apparent in the responses play a central role in determining people's conclusions, and hence are not mere rationalizations.

The real test of our position is whether a computer implementation of the theory produces the same conclusions as people do and for the same reasons, given the same information. In order to test out the core theory, we have built a computer model incorporating the reasoning patterns derived from our analysis (Baker, Burstein & Collins, 1987). Similar models were also built by Dontos and Zenankova (1988) and Kelly (1988) a student of Michalski. We plan to evaluate the theory in a series of experiments comparing the system's reasoning to that of expert human reasoners who have no special knowledge about the domain they are asked to reason about. To do this we will ask expert human reasoners, working from well-specified, small knowledge bases in geography and economics to draw plausible conclusions from each knowledge base and to estimate the certainty of each conclusion. The knowledge bases are incomplete and it is the subject's task to infer what they can about the missing information. For example, the geography data base has data about twelve different regions of the world concerning nine different variables such as climate, soil, terrain, precipitation, and grain.

This methodology for testing the theory looks like it will be very revealing. The data we have collected so far, though not fully analyzed as yet, are very rich in the kinds of inferences described in the core theory. But there are clearly new inference

patterns emerging in the data. We think this kind of tight coupling between computational modelling on the one hand and detailed analysis of human processing on the other hand offers a genuinely new approach to understanding human thinking.

7. GLOSSARY

argument. The concept within a statement to which a descriptor is applied. E.g., in "means−of−locomotion (birds)= {flying...}," "birds" is the argument.

argument transform. A plausible inference where a person infers a statement (or its negation) is true based on the fact that the argument in the statement is related by one of the four relations (GEN, SPEC, SIM, and DIS) to the argument in a statement the person believes is true. E.g., if a person believes that "grain(Kansas) = {wheat...}" and that "Iowa SIM Kansas" then a person may plausibly infer that "grain (Iowa) = {wheat...}". (See Tables 5 and 7 for other examples.)

certainty. The certainty parameter denoted by γ that indicates the degree of belief a person has that an expression is true. E.g., in "means−of−locomotion(dogs) = {swimming...}", γ denotes the degree of belief a person has that dogs in general can swim.

conditional likelihood. The certainty parameters denoted by α and β that in a mutual implication or dependency indicate the degree of constraint from one side of the expression to the other. E.g., "temperature(place) <−−−> desirability−of−living(place)", α denotes the degree to which "temperature(place)" predicts "desirability−of−living(place)" and β denotes the degree that "desirability−of−living(place)" predicts "temperature(place)".

dependency (see **mutual dependency**)

descriptor. The concept within a statement that applies to the argument to form a term. E.g., in "mean−of−locomotion(birds) = {flying...}", "means−of−locomotion" is the descriptor.

derivation from a mutual dependency. A plausible inference where a person derives a belief about a new statement based on knowledge about a particular statement and how another term depends on the term in that statement. E.g., if a person believes "temperature (place) <−$^+$−> desirability−of−living (place)", and that "temperature (Texas) = warm," then she may infer that "desirability−of−living (Texas) = high". (See Table 9 for other examples.)

44

derivation from mutual implication. A plausible inference where a person derives a belief about a new statement based upon knowledge about a particular statement and how another statement depends on that statement. E.g., if a person believes "grain(place) = rice <==> rainfall(place) = heavy" and that "grain(Southern China) = rice", then he may infer that "rainfall (Southern China) = heavy". (See Table 8 for other examples.)

DIS. The dissimilarity operator that specifies a concept that is dissimilar to another concept. E.g., geese DIS ducks in CX (birds, neck length(birds)) means that geese are dissimilar to ducks in the length of their necks.

dissimilarity transform. A plausible inference where a person infers a statement's negation based on the dissimilarity of the argument or referent in the statement to the argument or referent in another statement that the person believes is true. E.g., if a person believes that "geese DIS ducks in CX(birds, neck length (birds))" and that "sound (ducks) = quack" and that "neck length (birds) <--> sound (birds)", then she may conclude that "sound (geese) ≠ quack". (See Tables 5 and 7 for other examples.)

dominance. The certainty parameter denoted by δ that specifies the degree a subset comprises a large fraction of its superset. E.g., chicken comprise a large fraction of poultry, whereas turkeys comprise only a small fraction.

expression. Any statement, mutual dependency, or mutual implication. E.g., "sound(ducks) = quack", "geese DIS ducks in CX(birds, neck length(birds))", "cost(coal) <---> cost (oil)", "rainfall(place) = heavy <==> grain(place) = rice" are all expressions.

frequency. The certainty parameter, denoted by ϕ, that specifies the proportion of elements in the argument set for which the referent is true. E.g., for "means-of-locomotion (birds) = {flying..{", ϕ specifies the proportion of birds that fly.

GEN. The generalization relation that specifies a superordinate of a concept. E.g., "island GEN Great Britain in CX(islands, size (island))" means that Great Britain is an island (with typicality evaluated in terms of size).

45

generalization transform. A plausible inference where a person infers a statement is true based on the fact that the argument or referent in the statement is a generalization of the argument or referent in a statement the person believe is true. E.g., person knows that "birth-form (frogs) = eggs" and that "amphibians GEN frogs", the person might plausibly conclude that "birth-form (amphibians) = eggs". (See Tables 5 and 7 for other examples.)

implication. (see mutual implication)

multiplicity of the argument. The certainty parameter, denoted by μ_a, that specifies the degree to which there are multiple arguments within the superordinate of the argument for which the statement holds true. E.g., for "means-of-locomotion(birds) = {flying...}", μ_a is low because not many other kinds of animals can fly, whereas for "means-of-locomotion (robins) = {flying...}", μ_a is high because many other kinds of birds can fly.

multiplicity of the referent. The certainty parameter, denoted by μ_r, that specifies the degree to which there are multiple referents within the superordinate of the referent (i.e. the descriptor) for which the statement holds true. E.g. for "means-of-locomotion (birds) = {flying...}" μ_r is moderate because there are other means of locomotion (e.g. walking, swimming) among birds.

mutual dependency between terms. An expression which characterizes the relationship between two terms. E.g., "temperature (place) $<-^-->$ latitude (place)" expresses the relationship that temperature increases as latitude decreases. The α and β certainty parameters express the degree that knowing about temperature constrains latitude, and knowing about latitude constrains temperature, respectively.

mutual implication between statements. An expression which characterizes the relationship between two statements. E.g., "temperature(place)=hot $<==>$ latitude(place) = tropical" expresses the belief that hot places are tropical (the right arrow) and that tropical places are hot (the left arrow). The α and β certainty parameters express the degree to which knowing the place is hot leads to believing it is tropical, and the degree to which knowing the place is tropical leads to believing it is hot.

<u>referent</u>. The concept within a statement which specifies the value(s) of the term. E.g., in "means-of-locomotion(birds) = {flying...}", "flying" is the referent.

<u>referent transform</u>. A plausible inference where a person infers a statement (or its negation) is true based on the fact that the referent is related by one of the four relations (GEN, SPEC, SIM, and DIS) to the referent in a statement the person believes is true. E.g., if a person believes "political beliefs (George) = {conservative...}" and "hawkish SIM conservative", then she may plausibly conclude that "political beliefs (George) = {hawkish...}" (See Tables 5 and 7 for other examples).

<u>relation</u>. One of the four relations: GEN, SPEC, SIM, and DIS. They select a member of either the generalization set, the specialization set, the similarity set, or the dissimilarity set, respectively, of the set operated on. E.g. "birds GEN ducks" selects the set "birds" among the generalization sets of ducks, rather than water fowl or poultry.

<u>SIM</u>. The similarity relation that specifies a concept that is similar to another concept. E.g., geese SIM ducks in CX(birds; feet(birds)) means that geese are similar to ducks in the kind of feet they have.

<u>similarity</u>. The certainty parameter, denoted by σ, that specifies the degree of match between two concepts with respect to some set of characteristics specified by the context (CX). E.g., from "geese DIS ducks in CX(birds, neck length(birds))", σ specifies the degree to which ducks and geese are similar in the context of the neck lengths of birds.

<u>similarity transform</u>. A plausible inference where a person infers a statement is true based on the similarity of the argument or referent in the statement to the argument or referent in another statement the person believes is true. E.g., if a person believes that "geese SIM ducks in CX(birds, legs (birds))" and that "means-of-walking(ducks) = waddle" and that "legs(birds) <--> means-of-walking(birds)" then she may conclude that "means-of-walking (geese) = waddle". (See Tables 5 and 7 for other examples.)

<u>SPEC</u>. The specialization relation that specifies a subordinate of a concept. E.g., "bobolink SPEC bird in the CX (birds, characteristics (birds))" means a bobolink is a bird (with typicality evaluated in terms of all characteristics).

specialization transform. A plausible inference where a person infers a statement is true based on the fact that the argument or referent in the statement is a specialization of the argument or referent in a statement that the person believes is true. E.g., if a person believes "means–of–locomotion(birds) = {flying...}" and that "bobolink = SPEC(bird)" the person may plausibly conclude that "means–of–locomotion(bobolinks) = {flying...}". (See Tables 5 and 7 for other examples.)

statement. An expression where a descriptor is applied to an argument specifying some set of referents. E.g., "means–of–locomotion (birds) = {flying, hopping...}" is a statement.

term. The left side of a statement, that is, a descriptor applied to an argument. E.g., "means–of–locomotion (birds)" is a term.

transform based on a mutual dependency. A plausible inference where a person infers a dependency is true based on the fact that the argument in the dependency is related by one of three relations (GEN, SPEC, and SIM) to the argument in a dependency the person believes is true. E.g., if a person believes "latitude (place) <--> temperature (place)" and that "city SPEC place", then she can plausibly infer that "latitude (city) <--> temperature (city)" (See Table 13 for other examples).

transform based on a mutual implication. A plausible inference where a person infers an implication is true based on the fact that an argument or referent in the implication is related by one of the four relations (GEN, SPEC, SIM, or DIS) to the argument or referent in an implication the person believes is true. E.g., if a person believes that "means–of–locomotion(object)={flying...} <==> structural part (object) = {wings...}" and that "animals SPEC object", then he might plausibly infer that "means–of–locomotion (animals) = {flying...} <==> structural part (animal) = {wings...}" (See Tables 11 and 12 for other examples).

transitivity inference A plausible inference where a person infers that an implication or dependency is true by transitivity from the belief about two related implications or dependencies. E.g., if a person believes that "diet(person) = too much salt <==> blood-pressure (person) = too high" and that "foods-eaten (person) = processed foods <==> diet (person) = too much salt", then she may plausibly

conclude "foods-eaten (person) = processed foods <==> blood-pressure (person) = too high" (See Table 10 for other examples).

typicality   The certainty parameter, denoted by τ, that specifies the degree of match between a concept and its superordinate with respect to some set of characteristics specified by the context (CX). E.g., for "goose SPEC bird in CX (birds, neck length(birds))", τ denotes the degree that the neck length of geese is typical of the neck length of birds in general.

8. ACKNOWLEDGEMENTS

9. REFERENCES

Baker, M., Burstein, M.H., & Collins, A. (1987). Implementing a model of human plausible reasoning. In Proceedings of the Tenth International Joint Conference of Artificial Intelligence, Vol. 1 (pp. 185–188). Los Altos, CA: Morgan Kaufman.

Bobrow, D.G. & Winograd, T. (1977). An overview of KRL, a knowledge representation language. Cognitive Science, 1, 3–46.

Brachman, R.H. & Schmolze, J.G. (1985). An overview of the KL-ONE knowledge representation system. Cognitive Science, 9, 171–216.

Carbonell, J.R. & Collins, A., (1973). Natural semantics in artificial intelligence. Proceedings of Third International Joint Conference on Artificial Intelligence (pp. 344–351). Stanford CA: Stanford University.

Charniak, E. (1983). Passing markers: A theory of contextual influence in language comprehension. Cognitive Science, 7, 171–190.

Chomsky, N. (1965). Aspects of the Theory of Syntax. Cambridge, MA: MIT Press.

Cohen, P.R. (1985). Heuristic reasoning about uncertainty: An Artificial Intelligence approach. Boston: Pitman.

Cohen, P.R. & Grinberg, M.R. (1983). A theory of heuristic reasoning about uncertainty. AI Magazine, 4, 2, 17–24.

Collins, A. (1978a). Fragments of a theory of human plausible reasoning. In D. Waltz (Ed.) Theoretical Issues in Natural Language Processing II (pp. 194–201). Urbana, IL: University of Illinois.

Collins, A. (1978b). Human plausible reasoning Cambridge, MA: Bolt Beranek and Newman Inc., Report No. 3810.

Collins, A. (1985). Component models of physical systems. Proceedings of the Seventh Annual Conference of the Cognitive Science Society (pp. 80-89). Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins, A. & Gentner, D. (1982). Constructing runnable mental models. Proceedings of the Fourth Annual Conference of the Cognitive Science Society (pp. 86-89). Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins, A. & Gentner, D. (1983). Multiple models of evaporation process. Proceedings of the Fifth Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins, A. & Gentner D. (1987). How people construct mental models. In N. Quinn and D. Holland (Eds.) Cultural models in language and thought (pp. 243-265). New York: Cambridge University Press.

Collins, A.M. & Loftus, E.F. (1975). A spreading activation theory of semantic processing. Psychological Review. 82, 407-428.

Collins, A.M. & Quillian, M.R. (1972). How to make a language user. In E. Tulving & W. Donaldson (Eds.), Organization of Memory (pp. 309-351). New York: Academic Press.

Collins, A., Warnock, E.H., Aiello, N. & Miller, M.L. (1975). Reasoning from Incomplete Knowledge. in D. Bobrow & A. Collins (Eds.). Representation & understanding. Studies in cognitive science (pp. 383-415). New York: Academic Press.

Davies, T.R. & Russell, S.J. (1986). A logical approach to reasoning by analogy. Center for the Study of Language and Information. Stanford University.

Dontas, K. & Zemankova, M. (1987). APPLAUS: An Implementation of the Collins-Michalski Theory of Plausible Reasoning. Proceedings of the 3rd International Symposium on Methodologies for Intelligent Systems. Torino, Italy, 1988.

Gentner, D.G. & Collins, A. (1981). Studies of inference from lack of knowledge. Memory & Cognition, 9, 434-443.

Haviland, S.E. (1974). Nondeductive strategies in reasoning. Unpublished Ph.D. dissertation, Department of Psychology: Stanford University.

Johnson-Laird, P.N. (1980). Mental models in cognitive science. Cognitive Science, 4, 71-115.

Johnson-Laird, P.N. (1983). Mental models. Cambridge, MA: Harvard University Press.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press.

Kahneman, D. & Tversky, A. (1972). Subjective probability: A judgment of representativeness. Cognitive Psychology, 3, 430-454.

Kelly, J. (1988) PRS: A system for plausible reasoning. M.S. Thesis, Department of Computer Science, University of Illinois, Urbana.

Lukasiewicz, J. (1967). Many-valued systems of propositional logic. In S. McCall (Ed.), Polish logic. Oxford: Oxford University Press.

Martin-Lof (1982). Constructive mathematics and computer programming. In Methodology and Philosophy of Science VI. Amsterdam: North Holland Publishing Company.

Michalski, R.S. (1980). Pattern recognition as rule-guided inductive inference. IEEE Transactions on Pattern Analysis and Machine Intelligence. PAMI-2, 349-361.

Michalski, R.S. (1983). A theory and methodology of inductive learning. Artificial Intelligence, 20, 111-161.

Michalski, R.S. & Winston, P.H. (1986). Variable precision logic. Artificial Intelligence, 29, 121-146.

Minsky, M. (1975). A framework for representing knowledge. In P.H. Winston (Ed.), The psychology of computer vision (pp. 211–277). New York: McGraw–Hill.

Newell, A. (1980). Reasoning, problem solving and decision processes: The problem space as a fundamental category. In R. Nickerson (Ed.) Attention and Performance VIII (pp. 693–718). Hillsdale, NJ: Lawrence Erlbaum Associates.

Newell, A., & Simon, H.A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice–Hall.

Nilsson, N.J. (1986). Probabilistic logic. Artifical Intelligence, 28, 71–87.

Pearl, J. (1987). Embracing causality in formal reasoning. In Proceedings of the Sixth National Conference on Artificial Intelligence (pp. 369–373). Los Altos, CA: Morgan Kaufman.

Pearl, J. (1986). Fusion, propagation, and structuring in Bayesian networks. Artificial Intelligence, 29, 241–288.

Polya, G. (1968). Patterns of plausible inference. Princeton, NJ: Princeton University Press.

Quillian, M.R. (1968). Semantic memory. In M. Minsky (Ed.), Semantic information processing (pp. 216–270). Cambridge, MA: MIT Press.

Reiter, R. (1980). Logic for default reasoning. Artificial Intelligence, 13, 1–132.

Rips, L.J. (1975). Inductive judgements about natural categories. Journal of Verbal Learning and Verbal Behavior, 14, 665–681.

Rips, L.J. (1986). Mental muddles. In M. Brand & R.M. Harnish (Eds.), Problems in the representation of knowledge and belief. Tucson, AZ: University of Arizona Press.

Rips, L.J. (in press). Similarity, typicality, and categorization. In S. Vosniadou &

A. Ortony (Eds.). <u>Similarity and analogical reasoning</u>. New York: Cambridge University Press.

Rosch, E. (1975) Cognitive representations of semantic categories. <u>Journal of Experimental Psychology: General</u>, <u>104</u>, 192–233.

Salter, W. (1983). Tacit theories of economics. In <u>Proceedings of the Fifth Annual Conference of the Cognitive Science Society</u>. Rochester, NY: University of Rochester.

Schank, R. (1986). <u>Explanatory patterns: Understanding mechanically and creatively</u>. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Schank, R. & Abelson, R. (1977). <u>Scripts, plans, goals, and understanding</u>. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Shafer, G.A. (1976). <u>A mathematical theory of evidence.</u> Princeton, NJ: Princeton University Press.

Smith, E.E. & Medin, D.L. (1981). <u>Categories and concepts</u>. Cambridge, MA: Harvard University Press.

Stevens, A. & Collins, A. (1980). Multiple conceptual models of a complex system. In R. Snow, P. Federico, & W. Montague (Eds.), <u>Aptitude, learning and instruction: Cognitive process analysis</u> (pp. 177–197). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tversky, A. (1977). Features of similarity. <u>Psychological Review.</u> <u>84</u>. 327–352.

Tversky, A. & Kahneman, D. (1973). Availability. A heuristic for judging frequency and probability. <u>Cognitive Psychology,</u> <u>5</u>, 207–232.

Tversky, A. & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. <u>Progress in social psychology</u>, Hillsdale, NJ: Lawrence Erlbaum Associates.

Winograd. T. (1975). Frame representations and the declarative-procedural controversy. In D.G. Bobrow & A. Collins (Eds.). Representation and understanding (pp. 185–210). New York: Academic Press.

Woods, W.A. (1975). What's in a link. Foundations for semantic networks. In D.G. Bobrow and A. Collins (Eds.) Representations and understanding: Studies in Cognitive Science (pp. 35–82). New York: Academic Press.

Yager, R. (1987). Using approximate reasoning to represent default knowledge. Artificial Intelligence, 31. 99–112.

Zadeh, L.A. (1965). Fuzzy sets. Information and Control, 8, 338–353.