ARI Research Note 96-32

Testing and Refining a Core Theory of Human Plausible Reasoning

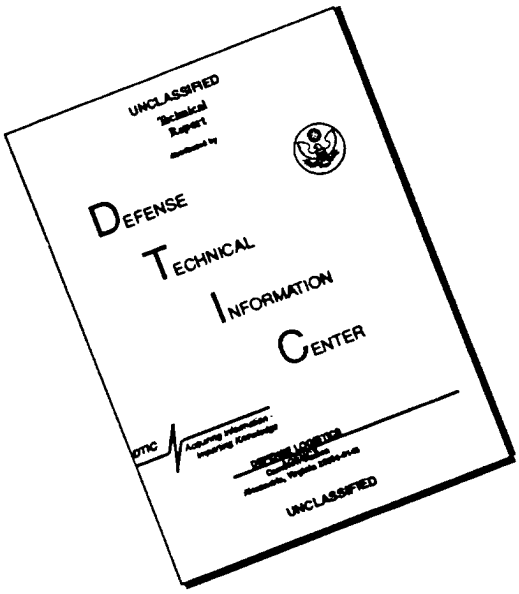
Allan Collins, Mark Burstein, and Michelle Baker
BBN Systems and Technologies Corporation

Research and Advanced Concepts Office
Michael Drillings, Acting Director

March 1996

United States Army
Research Institute for the Behavioral and Social Sciences

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Director

Research accomplished under contract
for the Department of the Army

Technical review by

Joseph Psotka

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED (from. . . to)
1996, March	Final	September 1985-September 1989

4. TITLE AND SUBTITLE	5a. CONTRACT OR GRANT NUMBER
Testing and Refining a Core Theory of Human Plausible Reasoning	MDA903-85-C-0411
	5b. PROGRAM ELEMENT NUMBER 0601102A
6. AUTHOR(S)	5c. PROJECT NUMBER B74F
Allan Collins, Mark Burstein, and Michelle Baker (BBN Systems and Technologies Corporation)	5d. TASK NUMBER 1019
	5e. WORK UNIT NUMBER C52

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
BBN Systems and Technologies Corporation 10 Moulton Street Cambridge, MA 02138	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. MONITOR ACRONYM
U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: PERI-BR 5001 Eisenhower Avenue Alexandria, VA 22333-5600	ARI
	11. MONITOR REPORT NUMBER Research Note 96-32

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

CORs: George Lawton and Judith Orasanu

14. ABSTRACT *(Maximum 200 words)*:

This report contains two papers prepared during the last year on the contract. The first paper details our extensions of a formal theory of human plausible reasoning, and the second paper is an overview of the theory and experimental work that appeared as a book chapter.

15. SUBJECT TERMS

Reasoning Information processing Artificial intelligence Cognitive psychology

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	80	

Testing and Refining a Core Theory of Human Plausible Reasoning

Allan Collins Mark Burstein
Bolt Beranek and Newman Inc.
Cambridge, MA

Michelle Baker
Columbia University
New York, NY

cattle in the Chaco. This protocol also illustrates how people combine evidence from different plausible inferences to reach a final conclusion.

The second protocol illustrates a plausible deduction. It is from a series of questions we asked different respondents (Collins 1978).

Protocol 2

Q. Is Uruguay in the Andes Mountains?

A. 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 subject is making a plausible deduction called a specialization transform in the theory. He thinks that the Andes mountains are in most South American countries, so they are likely to be in Uruguay. There are two certainty parameters in the theory that show up here: the higher the frequency of countries that have the Andes, and the more typical Uruguay is, the more certain the inference.

1.1 The Core Theory of Collins and Michalski

There are four types of expressions in the core theory of Collins and Michalski (1989) that are shown in Table 1. The first are simple statements consisting of a descriptor **d** (e.g. means-of-locomotion) applied to an argument **a** (e.g. birds) and realized by a referent **r** (e.g. flying). The brackets and dots around the referent indicate that there may be other means of locomotion for birds, such as walking. The second kind of expression involves one of four relations: GEN for generalization, SPEC for specialization, SIM for similarity, and DIS for dissimilarity. Each relational statement specifies a context (CX) where the first variable is the domain over which typicality or similarity are computed, and the second variable

Table 1

Different Types of Expressions in the Core Theory

Statements (S)

$d(a) = r$

means-of-locomotion(birds)={flying...}

Relational Statements (R)

a_1 REL a_2 in CX (A,d) where REL = GEN, SPEC, SIM, or DIS

bird GEN robin in CX (birds, all characteristics)

chicken SPEC fowl in CX (birds, biological characteristics)

duck SIM goose in CX (birds, habitat)

duck DIS goose in CX (birds, neck length)

Mutual Implications (I)

$d_1(a) = r_1 <=> d_2(a) = r_2$

temperature(place)=warm & rainfall(place)=heavy <=> grain(place)=rice

Mutual Dependencies (D)

$d_1(a) <--\pm--> d_2(a)$

average temperature (place) <--~--> latitude (place)

is the descriptor(s) with respect to which typicality or similarity are computed. The last two examples of relational statements represent the fact that ducks and geese are similar in their habitats, but dissimilar in neck length.

The other two types of expressions in Table 1 are called mutual implications and mutual dependencies. A mutual implication specifies how one statement is related to another statement. The example states that warm temperature and heavy rainfall imply rice growing, and vice versa. A mutual dependency relates two terms (e.g. latitude (place) and temperature (place)). The example represents the belief that the average temperature of a place is inversely related to its latitude.

Table 2 shows a pattern of eight statement transforms from the core theory (Collins & Michalski, 1989). Given a person believes that the flowers of England include daffodils and roses, the first four transforms all vary the argument, England. Given no other information, it is a plausible inference that daffodils and roses are flowers of Europe in general (a generalization transform). Also, it is a plausible inference that Surrey, which is a small county in England, has daffodils and roses (a specialization transform); that Holland, which is similar to England in its climate, has daffodils and roses (a similarity transform); and that Java, which is quite dissimilar to England in climate, does not have daffodils and roses (a dissimilarity transform).

The other four transforms vary the referent, daffodils and roses. If you believe that daffodils and roses are flowers of England, it is plausible that most temperate flowers grow there (a generalization transform), that yellow roses grow there (a specialization transform), that peonies grow there (a similarity transform), and that bougainvillea, a tropical plant, does not grow there (a dissimiliarity transform). These eight transforms were one of four classes of plausible inference in the core theory.

Table 2

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

Argument-based 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(Java)={daffodils, roses...}

Referent-based 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...}

Table 3 shows how different parameters affect the certainty of these eight plausible inferences:

•Typicality (τ) affects GEN and SPEC transforms. The more typical England is of Europe, or Surrey is of England, with respect to climate (or any variable that affects flower growing), the more certain the inference.

•Similarity (σ) affects the SIM and DIS transforms. Hence, the more similar Holland is to England, and the less similar Java is to England, with respect to climate, the more certain the inference.

•Conditional Likelihood (α) reflects the degree to which climate (or any variable that affects flower growing) determines what flowers are grown in a place. The more effect climate has on flower growing, the more certain any of these inferences.

•Frequency (ϕ) reflects the all/some distinction in logic, but as a continuous variable. When applied to an instance like England, frequency only makes sense if it is the frequency of daffodils and roses in different parts of England. The more frequent daffodils and roses are in England, the more likely they are found in Europe, Surrey, Holland, or even Java.

•Dominance (∂) applies to GEN and SPEC inferences and reflects the degree the subset comprises a large part of the set. For example, since Surrey in only a small part of England, the inference about growing daffodils and roses is less certain than for Southern England as a whole.

•Multiplicity of the argument (μ_a) reflects the degree to which more than one country (the superordinate of the argument) has daffodils and roses. Since many countries presumably have daffodils and

Table 3

Effects of Different Parameters on Statement Transforms

Transforms in Table 2		Parameters							Target Node
		τ	σ	α	ϕ	δ	μ_a	μ_r	
Argument-Based	1 GEN	+	0	+	+	+	+	0	Europe
	2 SPEC	+	0	+	+	+	0	0	Surrey
	3 SIM	0	+	+	+	0	+	0	Holland
	4 DIS	0	-	+	-	0	-	0	Brazil
Reference-Based	5 GEN	+	0	+	+	+	0	+	Tropical Plants
	6 SPEC	+	0	+	+	+	0	0	Yellow Roses
	7 SIM	0	+	+	+	0	0	+	Peonies
	8 DIS	0	-	+	-	0	0	-	Bougainvillea

+ means higher values of parameter increase the certainty of the inference, and - means higher values of parameter decrease the certainty of the inference.

roses, μ_a is high and the argument-based inferences are more certain (except for the DIS inference).

•Multiplicity of the referent (μ_r) reflects the degree to which England has more types of flowers (the superordinate of the referent) than daffodils and roses. Since countries usually have many different types of flowers, μ_r is high and the referent-based inferences more certain (except for the DIS inference).

In addition to these seven parameters, the certainty of each of these inferences is affected by the certainty (γ) of the person's belief in each of the premises in the inference. For example, the more certain the person is that England produces daffodils and roses, and that flowers depend on climate, the more certain the inference. These various parameters are described in more detail in the earlier paper (Collins & Michalski, 1989).

Table 4 shows how two of the inferences shown previously are represented formally in the theory. The first shows the similarity transform from Protocol 1 where the tutor inferred that cattle might be raised in the Chaco, because it was similar to western Texas. He must have been certain that cattle were raised in Texas (γ=high), that cattle are raised in many places other than Texas (μ_a=high), and that different parts of western Texas have cattle (ϕ=high). He seemed to think that Chaco was at least moderately similar (σ=moderate) to western Texas with respect to variables, such as vegetation, that determine whether a place can support cattle raising (α=moderate likelihood). He seemed to derive only fairly low certainty (γ=low) from this inference that cattle might be raised in the Chaco.

The second inference shown is from Protocol 2 where the respondent inferred that the Andes might be in Uruguay. He thought that the Andes were in most South American countries, so frequency (ϕ) was at least moderate, and his certainty (γ) about that was fairly high. He knew Uruguay was a very typical South American country in most respects

Table 4

Examples of Formal Representation in the Core Theory

Similarity Transform from Protocol 1

livestock (Western Texas) = cattle: γ = high, μ_a = high, ϕ = high
Chaco SIM Western Texas in CX(region, vegetation): γ = moderate, σ = moderate
vegetation (region) <----> livestock (region): α = moderate, γ = high
Chaco, Western Texas SPEC region: γ = certain
livestock (Chaco) = cattle: γ = moderate

Specialization Transform from Protocol 2

mountains (South American country) = Andes: ϕ = moderate, γ = high
Uruguay SPEC South American country in CX(country, all characteristics): τ=high
characteristics (country) <-----> mountains (country): α = low, γ = high
Uruguay, South American country SPEC country: γ = certain
mountains (Uruguay) = Andes: γ = moderate

(τ=high), but that has only a weak relation to whether a particular mountain range is there (α=low). So he concluded with moderate certainty that the Andes were in Uruguay.

There were three other classes of plausible inferences in the core theory developed by Collins and Michalski (1989), which are exemplified in Table 5. First, there were derivations from implications and dependencies. For example, if a person believes warm places with heavy rainfall produce rice and that the Amazon region is warm and has heavy rainfall, one might infer that they probably grow rice in the Amazon. Second, there were transitivity inferences on implications and dependencies. For example, if one believes that the humidity of a place is directly related to its average temperature, and that the average temperature of a place is inversely related to its latitude, then one might plausibly infer that humidity of a place is inversely related to its latitude. Third, there were transforms on implications and dependencies. For example, if one believes that places with a subtropical climate produce oranges, then one might infer that they produce other citrus fruits as well. The different variants of these three classes of inference are detailed in Collins and Michalski (1989).

This summarizes the core theory developed earlier. Subsequent to the development of the core theory, we ran an experiment using a technique developed by Michelle Baker, described in section 2. Also we systematically examined the space of all possible inferences that could be generated given the different kinds of expressions in the core theory. These two efforts have led to a revised theory which we outline in section 3.

Table 5
Examples of Other Classes of Plausible Inferences in the Core Theory

Derivations from Implications and Dependencies

temperature(place)=warm & rainfall(place)=heavy <=> grain(place)=rice
temperature(Amazon) = warm
rainfall(Amazon) = heavy
Amazon SPEC place
grain(Amazon) = rice

Transitivity Inferences on Implications and Dependencies

humidity (place) <--$^{+}$--> average temperature (place)
average temperature (place) <--$^{-}$--> latitude (place)

---

humidity (place) <--$^{-}$--> latitude (place)

Transforms on Implications and Dependencies

climate (place) = subtropical <=> fruit (place) = {oranges...}
citrus fruit GEN oranges in CX (fruit, growing conditions)
growing conditions (fruit) <-----> place (fruit)

---

climate (place) = subtropical <=> fruit (place) = {citrus fruit...}

2. An Experiment on Human Plausible Reasoning

In order to test whether the core theory encompassed all the different plausible inferences that expert reasoners would make given a partial database of facts, such as that developed for our simulation of the core theory (Baker, Burstein, and Collins, 1987; Burstein and Collins, 1988), we sought a method for collecting protocols of plausible inferences from a consistent set of facts. To this end, Michelle Baker developed a matrix of different interrelated variables in geography, crossed by countries, as shown in Table 6. She then interviewed five different scientists (who were not geographers) as they attempted to fill in the missing cells in the matrix.

The experimental procedure for each subject was as follows. Subjects were first presented with a graph showing the set of variables from the matrix and some unlabeled and undirected links representing possible dependencies between the variables. They were asked to put arrowheads on the links to indicate the direction of cause and effect, and label the arrows with their estimates of the strength of effect (α and β) of the variables on each other. Figure 1 shows one subject's dependency network filled out. Some subjects added new variables, such as vegetation, in discussing their own understanding of the interdependencies of the domain.

Subjects were then shown the entire matrix and asked to fill in the missing cells, in whatever order they chose. They were asked to verbalize their reasoning as they tried to fill in each cell, and were prompted to expand on that reasoning anytime the reasoning was unclear. The sessions varied in length from one half hour to one and one half hours for different subjects. Each session was recorded on audio tape and transcribed.

In analyzing the transcripts of these sessions, we sought to identify and formalize as many of the plausible inferences as we could find, each time considering whether the formal theory, as described in Collins and

TABLE 6

Matrix Used in Experiment

12 x 9	Climate	Water Supply	Grain Grown	Has River?	Precip-itation	Season Description	Soil Type	Temperature Range	Terrain
Afghanistan	?	?	NONE	?	?	?	Brown Grey	Hot Very Hot	Mountains
Angola	?	Moderate Abundant	Corn	YES	Abundant	Summer Rain	Dark Brown Grey	Hot	?
Egypt	Dry Climate	Moderate (Irrigated)	Wheat	YES	Very Light	?	Grey	Very Hot	Plains
Florida	Subtropical Humid Trop.	?	Corn	?	Moderate Abundant	Mild Winter Long Summer Even Rain	?	?	Lowlands Plains
Iran	Semi-Arid Mediteranean	?	?	NO	Light	Winter Rain	Grey	?	?
Italy	Mediteranean	Moderate	?	YES	?	Mild Winter Hot Summer Winter Rain	Complex Red-Yellow	Mild Hot	Mountains Plains
Java	Humid Tropics	?	Rice Corn	NO	Abundant Very Wet	No Winter Even Rainfall	?	Hot	Mountains Lowlands
Louisiana	Subtropical	Abundant	?	YES	?	Mild Winter Long Summer Even Rainfall	Red-Yellow Black	?	Lowlands Plains
Peru	Highland Arid	Moderate (Irrigated)	Corn Rice	?	Very Light Light	Summer Rain	Complex	?	Mountains
Saskatchewan	Dry Climate	?	Wheat Oats,Rye	YES	Light	Winter Rain	Dark Brown Brown Complex	Cool Mild	Plateau
Upper Volta	?	Abundant	Rice Millet	YES	Very Wet	?	?	Hot Very Hot	Lowlands Plains
West Indies	Humid Tropics	Abundant	Rice Corn	NO	Abundant Very Wet	No Winter Even Rainfall	Red-Yellow	Hot	?

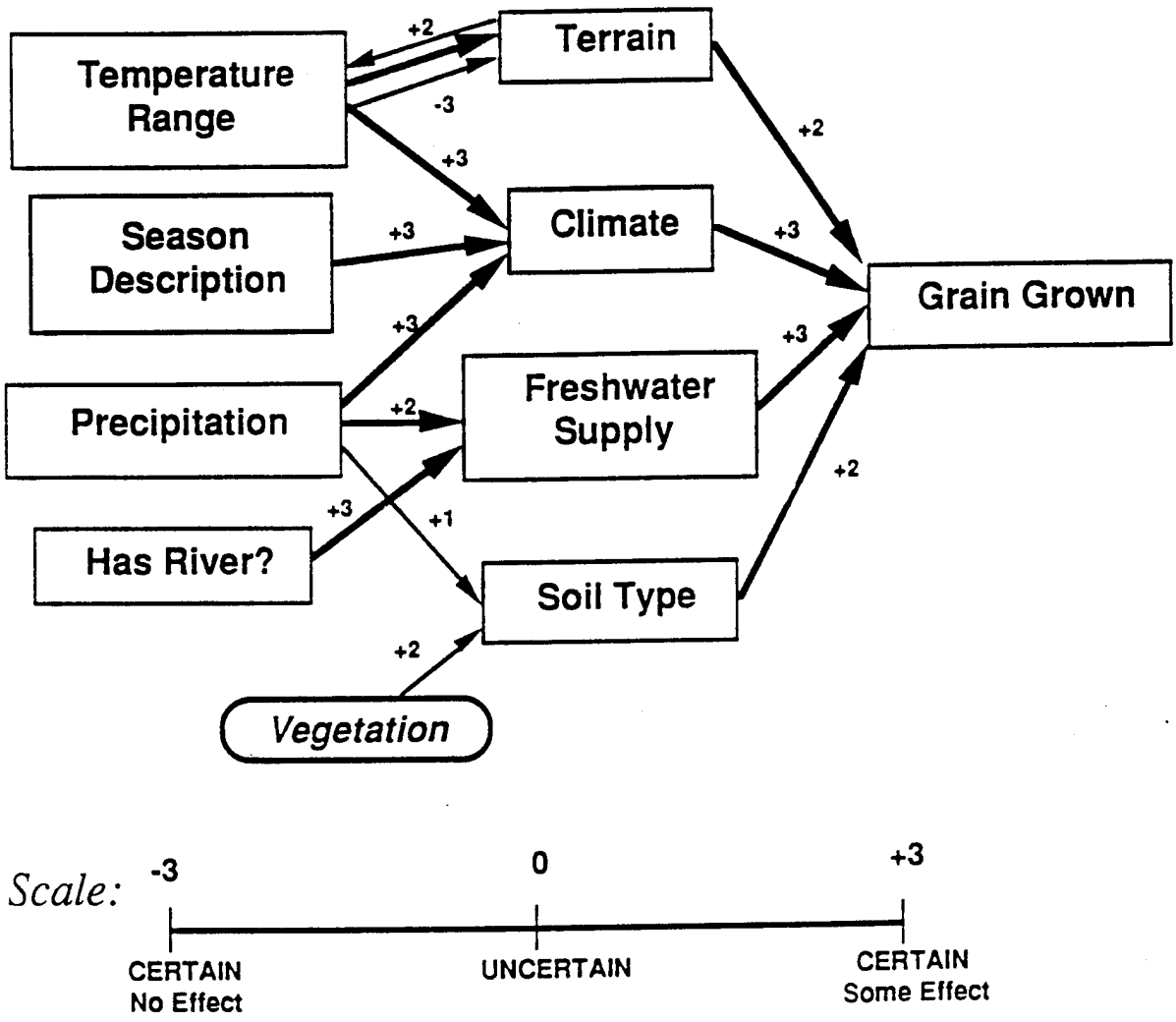


Figure 1. One subject's dependency network. (Thick arrows an boxes were given; thin arrows, arrowheads, scales, and the rounc box were added by the subject.)

Michalski (1989) accounted for the observed behavior. Where there were discrepancies, we considered how the theory was inadequate, and considered possible extensions and their ramifications. This section discusses some selected samples of these protocols and the issues they raised.

2.1 A Sample Protocol

It became clear in analyzing the protocols that there were a number of inference patterns that our experimental approach tended to highlight. Due to the tabular presentation of the information about different countries, it was as easy to reason *backwards* from things normally considered effects as it was to reason *forward*, to plausibly infer values for the variables in those effects from the variables that they depended on. It also made it easier for people to correlate a number of variables for several different places when reasoning in both the forward and backward directions.

One example of an issue that arises in uncertain abductive reasoning from effects to causes is found in the protocol of a subject (Subject 1) who tried to use the value for the amount of available fresh water supply in Italy to reason backward to infer the value for precipitation in Italy. In the matrix, there were two variables that directly affected water supply. The principal one was a qualitative value (light, moderate, abundant) for the average amount of precipitation of the country. The second was a variable indicating whether there were any rivers in the country or not (yes or no). For Italy, the water supply was listed as being moderate, and the column labeled HAS-RIVER? had value YES.

<u>Protocol 3</u>

S1: Let's go back and do Italy first then... What the mountains tell you is that increases the precipitation. And the Mediterranean climate tells you that it doesn't typically have a lot; Mediterraneans tend to be fairly dry climates. So my guess about Italy is that it probably... but the fresh water supply also implies... well it could get its fresh water

all from the rivers, so the moderate fresh water supply... because with Egypt had moderate and that other one I inferred was moderate. My inclination would be to say that implies that there is not a lot of rainfall, okay. But the mountains imply that there is rainfall, okay. So that leads me to... I'm not sure what variables I have for rainfall, very light and light, so I'd go for light.

There are several inferences taking place here. In the first part of this response, the subject focused on the evidence that the Italian climate was Mediterranean, and the fact that there were mountains. The Mediterranean climate led the subject to infer that Italy had limited precipitation, while the presence of mountains indicated that there would be more rain than other, similar, lowland areas with the same general climate. Both of these inferences are based on *implications*, although the second one requires an extension to the theory that we will come back to shortly.

In the second half of the response, the subject based his inferences on the evidence that the fresh water supply given for Italy was moderate, and the fact that there were rivers. As described earlier, both rivers and precipitation are contributing factors to water supply. There are two kinds of uncertain information here. One is the question of how much of a contribution each of those factors can make to overall water supply, a question for which the subject presumably had little direct knowledge. The other problem is the lack of information of even a qualitative kind as to the amount of water available from rivers. All the matrix supplied was the fact that there were at least *some* rivers.

It appears from the pattern of this subject's protocols, and from subsequent questioning of the subject, that he normally treated water supply as if it was directly dependent on precipitation, independently of the presence of rivers. In general, either precipitation or rivers could account for the water supply of a place, and this subject generally assumed water supply was directly correlated with precipitation, unless there was evidence to the contrary. This led to the incomplete comment "but the fresh water supply also implies...". Later questioning confirmed that he

was starting to say that the moderate fresh water supply indicated moderate precipitation, a backward, abductive inference from the dependency. This was quickly followed by the comparison that Egypt had moderate water supply even though the precipitation there was very light, because there was a river. Thus the analogy to Egypt supported the conclusion that the precipitation was **very light.** By combining evidence from three sources: the analogy to Egypt, the presence of mountains, and the Mediterranean climate, the subject concluded that the precipitation was probably **light.**

This reasoning is formalized as follows:

Terrain(place) = mountains <==> Precipitation(place) > *"normal"*
Terrain(Italy) = mountains

(1*) Precipitation(Italy) > *"normal"*

Climate(place) = Mediterranean <==> Precipitation(place) = light
Climate(Italy) = Mediterranean

(2) Precipitation(Italy) = light

Precipitation(place) <--+--> Water-supply(place)
Water-supply(Italy) = moderate

(3) Precipitation(Italy) = moderate

Precipitation(place) <--+--> Water-supply(place)
Has-rivers(place) <--+--> Water-supply(place)
Water-supply(Italy) = moderate
Has-rivers(Italy) = yes

(4*) Precipitation(Italy) ≠ moderate *(Discount 3)*

Water-supply(Egypt) = moderate
Water-supply(Italy) = moderate
Has-rivers(Egypt) = yes
Has-rivers(Italy) = yes

(5) Italy SIM Egypt in CX(countries, Has-rivers & Water-supply)

Precipitation(place) OR Has-rivers(place) <--+--> Water-supply(place)
Italy SIM Egypt in CX(countries, Has-rivers & Water-supply) *(from 5)*
Precipitation(Egypt) = very light

(6) Precipitation(Italy) = very light

Precipitation(Italy) > *"normal"* *(from 1)*
Precipitation(Italy) = light *(from 2)*
Precipitation(Italy) = very light *(from 6)*

(7*) Precipitation(Italy) = light

By this analysis, there are several issues raised in the protocol that are not specifically covered by the core theory (indicated by *). The first is the use in inferences 1 and 7 of inequalities rather than equal signs. The consideration of the general issue of continuous variables and inequalities will be taken up later in this and the next sections.

The first inference (1*) also raises an issue for the Collins and Michalski (1989) theory that is implicit in the use of default values in Minsky's (1975) frame paper. This inference was a kind of reasoning based on a norm or default value. The logic of the reasoning is this: Whatever is determined to be the normal value of rainfall in a place based on other variables, mountains tend to make the rainfall higher. So if Mediterranean climates have light rainfall, the mountains would make the rainfall greater than light. "Normal" is a "dummy value" for the precipitation variable used in order to carry forward the reasoning. This dummy value is filled in by two other inferences (2 and 6) and the average value computed from those inferences is adjusted upward in 7 to incorporate the adjustment specified in (1*).

A third problem for the core theory occurs in inference 4, where "counter evidence" to inference 3 is considered. This inference type has been called a *Functional Alternative* meta-inference in Collins (1978), and

is quite common. The pattern occurs when there are several variables that independently can influence a dependent variable. This can either be written as:

$$d_1(a) \text{ OR } d_2(a) \text{ <---> } d_3(a)$$

or, equivalently, but staying within the syntax of the original core theory, as two separate dependencies:

$$d_1(a) \text{ <---> } d_3(a) \text{ and } d_2(a) \text{ <---> } d_3(a)$$

Suppose an inference has been made from a dependency, to infer a value for the independent variables as by:

$$\begin{array}{ll} d_1(a) \text{ <---> } d_3(a) & \\ \underline{d_3(a) = r} & \text{for } r \in \{\text{high, medium, low}\} \\ d_1(a) = r & \end{array}$$

Then suppose independent evidence shows that $d_2(a) = r$, accounting for $d_3(a) = r$ by different means. By a *Functional Alternative* meta-inference, this invalidates or drastically reduces the certainty of the original inference that concluded $d_1(a) = r$. Thus, in the protocol, when it is discovered that Italy has rivers, and this accounts for Italy's moderate water supply (by analogy to Egypt), then that decreases the certainty of inference 3 that the Italy's moderate water supply implies that it has moderate precipitation. This rule is essentially an application of Occam's razor to plausible inferences with dependencies. It does not constitute evidence that the original inference was wrong, just that the evidence used to make the inference can be accounted for by other means. The set of meta-inferences is described most fully in Collins (1978). They are not included in the formalized core theory of Collins and Michalski (1989), so a full treatment of this and the other meta-inferences observed in the protocols is still an open problem.

The fourth issue raised by this protocol that requires an extension of the core theory occurs in inference 6. Informally, this is an extension of the inference pattern for reasoning *backwards* with a dependency, to the

case where there are multiple dependencies active. For the particular case of inference 6 there are two things affecting the water supply of a place, precipitation and the presence of rivers. The subject sees that there are two places (Italy, Egypt) which are similar on the affected variable, Water-supply, and one of the independent variables, Has-rivers. He plausibly concludes that they are similar on the other independent variable Precipitation. The extended rules for these dependency inferences are shown in Section 3.

2.2 Examples of Plausible Generalization

The next portion of the protocol of Subject 1 illustrates a new component of the theory, the formation of a *generalization* (in this case, the formation of a new implication) from the water supply variables for Italy and Egypt, and how that knowledge is used to guide his inferences about Louisiana. In the matrix, Louisiana was given as having a subtropical climate, abundant water supply, rivers, and a terrain of lowlands and plains.

Protocol 4

S1: Louisiana ... Precipitation, what is the precipitation? So the places with just a river and very little rainfall were moderate in their fresh water supply, and this is abundant. Now, unfortunately that is a case where I really know that Louisiana has a lot of rainfall. But that would be the nature of my inference, that it at least has a moderate precipitation ... from the fresh water supply.

This protocol reveals that sometime between the earlier protocol, where he reasoned about Egypt and Italy, and this one, he made a generalization that what was true of Egypt and Italy was true of all places. The generalization from Egypt and Italy is formalized as follows:

Has-river(place) & Precipitation(place) <---> Water-supply(place)

Has-river(Egypt) = yes

Precipitation(Egypt) = very light

Water-supply(Egypt) = moderate

Has-river(Italy) = yes

Precipitation(Italy) = light *(by inference 7 above)*

Water-supply(Italy) = moderate

(8) Has-river(place) = yes & Precipitation(place) = light

$\Longleftrightarrow$ Water-supply(place) = moderate

This generalization is one of the new rules described in Section 3, a generalization to form an implication from a dependency and data about a particular case (two similar cases in this instance, although the data about the precipitation level for Italy was inferred on the basis of the single example of Egypt). Generalizations like inference 8, where an existing dependency is combined with a specific example to form an intermediate statement (the implication), are essentially the analogs in our plausible reasoning theory of the "chunking" process in SOAR (Laird, Rosenbloom & Newell, 1986), and *explanation-based generalizations* as described by DeJong (1981) and Mitchell (1983). All of these generalization mechanisms hinge on the combination of general causal or explanatory background knowledge with a new specific case or cases to form a new, potentially more useful general rule. We call the class of such generalizations *refinements*.

2.3 Reasoning with Inequalities

Once the generalization just described had been formed, the protocol given above shows how it was used. The inference is basically that places with rivers and a little rainfall have moderate water supply, so places with abundant water supply must get more rain. Louisiana's abundant water supply, being greater than both Egypt and Italy's, means that it should have greater precipitation as well.

Has-river(place) & Precipitation(place) <--+--> Water-supply(place)

Has-river(place) = yes & Precipitation(place) = light

<==> Water-supply(place) = moderate

Has-river(Louisiana) = yes

Water-supply(Louisiana) = abundant

(9) Precipitation(Louisiana) > light (or ≥ moderate)

In the protocol, the subject reached the conclusion that since Louisiana had an abundant water supply, then it must have "at least moderate" precipitation. In the formalization of this inference, we have described the pattern as predicting simply that it should be "greater than light", light being the corresponding value in the implication. We take these as equivalent with respect to a {low, medium, high} scale. (For precipitation, the term **light** corresponds to the more neutral referent **low**, and **abundant** is the same as **high**.)

This is an example of a whole class of inferences that were not explicitly dealt with in the original core theory. The issue is one of reasoning with inequalities on continuous or ordered variables, in conjunction with dependencies between those types of variables. These inferences all depend on the presence of a *specified* dependency: specified dependencies are those labeled with a + or - to indicate that an increase/decrease in the values for a term on one side has a corresponding positive or negative effect on the other. We will deal with this issue more fully in Section 3.

2.4 Reasoning Using Multiple Dependent Factors

When it came time to consider which grains could be grown in different places, subjects were faced with a situation where all of the descriptors in the matrix were potentially relevant to some degree. Consider again the dependency graph of Figure 1. By this graph (and, of course, belief in this set of dependencies varied from subject to subject), water supply, climate, soil type and terrain all directly influence what

grains are grown, and the other variables in the chart contribute to those four.

The pattern that emerged when subjects looked to make inferences where a number of contributing factors were known was quite consistent. They first tried to form generalizations, and then used those generalizations to answer the questions. These generalization inferences occurred quite frequently, and not always when they helped answer a specific question. At a procedural or strategic level, we have observed two different paths to the formation of generalizations from multiple factors. The first is a generalization strategy based on a dependency and two similar examples. This is really a form of guided induction. Protocol 5, below is a clear example of this. The second strategy observed was to form a weak generalization based on a single example and a dependency, as described in Section 2.2, and then refine that generalization as new confirming examples were encountered. Protocol 6 gives an example of that strategy.

We look first at an example from Subject 1, when he was looking for a way to answer what the terrain of the West Indies was.

Protocol 5

S1: In the West Indies I'm up to and its terrain.. I don't have any good terrain inferences. Humid tropics. Red and Yellow [soil]. I can't infer. So, here we have another humid tropic with rice and corn and we had one of those in Java. *So humid, tropic climates seem to be leading to rice and corn and abundant wet precipitation.*

Subject 1 is making two inferences here. One is that humid, tropical climates determine abundant precipitation, which is almost by definition. The other is the generalization that these factors determine the grains grown to be rice and corn. We formalize the latter inference below. Since it is common knowledge that humid, tropical places are hot and wet, our formalization of this inference and the next several to come include the

temperature, precipitation and water supply factors as part of the generalizations made.

The generalization about rice and corn in this protocol is formalized as follows:

Climate(place) & Water-supply(place) & Precipitation(place) &
Season-description(place) & Soil-type(place) &
Temperature-range(place) & Terrain(place)
<---> Grain-grown(place)

Climate(Java) = humid tropics
Precipitation(Java) = abundant
Temp-range(Java) = hot
Grain-grown(Java) = rice, corn
Climate(West Indies) = humid tropics
Precipitation(West Indies) = abundant
Temp-range(West Indies) = hot
Grain-grown(West Indies) = rice, corn

(10) Climate(place) = humid tropics
& Precipitation(place) = abundant & Temp-range(place) = hot
<==> Grain-grown(place) = rice, corn

Later, this generalization was refined when Subject 1 came to the following conclusions about the grains grown in Louisiana (we include a portion of the matrix for comparison, with values in *italics* that the subject had filled in previously):

	Climate	Water Supply	Precip-itation	Season Descrip.	Temp. Range	Terrain	Grain Grown
Java:	Humid tropics	***Abund.***	Abund.	No Winter Even Rain	Hot	Mountains Lowlands	Rice Corn
West Indies:	Humid tropics	Abund.	Abund. Very Wet	No Winter Even Rain	Hot	?	Rice Corn
Upper Volta:	***Humid tropics***	Abund. Wet	Very	***No Winter Even Rain***	Hot Very Hot	Lowlands Plains	Rice Millet
Florida:	Subtrop. Hum. trop.	***Moder. Abun.***	Moderate Abund.	Mild Winter Long Summer Even Rain	?	?	Corn
Louisiana:	Subtrop.	Abund.	***Abund.***	Mild Winter Long Summer Even Rain	***Hot***	Lowlands Plains	?

Protocol 6

S1: So, grain grown: Mild winter, long hot summer, rainfall evenly distributed. We have abundant fresh water supply. Subtropical climate we know. The lowlands are plains. So my picture that I'm getting is a little bit corn and rice, but because of the abundant rainfall, these (Java, Upper Volta) are both rices, and this (Florida) is a corn. These are rice, this (Florida) is a [more] moderate climate. I inferred that Java must have abundant fresh water supply *so this warm climate, rainfall evenly distributed, and abundant water pattern seems to go with rice and corn*, so it looks like rice and corn.

E: What certainty is that?

S1: That's only moderate certainty. I mean even wheat.. There was a place where ... on millet I don't have enough information. Wheat was this dry climate thing and I think I probably... See, all these variables interact, so the better the pattern, the better the fit.

There are several inferences here, again based on the formation of *generalizations on several variables* relating to the question being investigated. The subject is trying to decide which factors, similar to those present in Louisiana, correlate with different grains. The implication developed in Protocol 5 about rice and corn is reiterated and refined here in two different ways to conclude that both rice and corn are grown in hot places with abundant water supply and even rain. There is also evidence that a generalization was made earlier that wheat was a dry climate crop. The subject declined to make a generalization about millet because there was only one place in the matrix that listed it. The theory would certainly have sanctioned a similarity-based inference that Louisiana could have millet, but the subject apparently declined to make it, probably because he was too uncertain about the conclusion.

Taking these generalizations in order, the first inference in this protocol is the refinement of the generalization made earlier about rice and corn. This seemed to occur in two steps. One used Java and Upper Volta as examples to conclude that rice is grown in hot places with abundant water supply; the other used Java and Florida to conclude that corn is grown in humid tropical and subtropical places with abundant precipitation and a similar set of season descriptions (mild winter, long summer, even rainfall). Although Java was used in the original generalization, we assume that it was reused in these refinements both because of the experimental condition that the subject was looking at the matrix while answering, and because the generalization refinement introduces new factors that were not in the earlier generalization.

In formalizing this implication refinement, we have noted places where several referents were generalized together in the implication by placing them in brackets ($[r_1, r_2,...]$), in the order of their believed frequency (ϕ) based on the evidence used. This is consistently done for implication, dependency and statement generalization, as described in section 3. Thus, in the resulting implication, the temperature range of hot dominates, as does a lowlands terrain, and rice is the most strongly

predicted grain, with millet only weakly predicted. First we look at the rice case:

Climate(place) & Water-supply(place) & Precipitation(place) &
Season-description(place) & Soil-type(place) &
Temperature-range(place) & Terrain(place)
<---> Grain-grown(place)

Climate(place) = humid tropics &
Precipitation(place) = abundant & Temp-range(place) = hot
<==> Grain-grown(place) = rice, corn *(from 10)*

Climate(Java) = humid tropics
Water-supply(Java) = abundant
Precipitation(Java) = abundant
Temp-range(Java) = hot
Season-description (Java) = no winter, even rain
Grain-grown(Java) = rice, corn
Climate(Upper Volta) = humid tropics
Water-supply(Upper Volta) = abundant
Precipitation(Upper Volta) = abundant
Temp-range(Upper Volta) = hot, very hot
Season-description(Upper Volta) = no winter, even rain
Grain-grown(Upper Volta) = rice, millet

(11) Climate(place) = humid tropics &
Water-supply(place)=abundant & Precipitation(place) = abundant
& Temp-range(place)= [hot, very hot]
& Season-description(place) = no winter, even rain
<==> Grain-grown(place)=[rice (ϕ = 1) , corn (ϕ = .75), millet (ϕ = .25)]

The second generalization refinement used Florida, essentially to increase the certainty that corn was grown in these same hot, wet places. This case also raises the possibility of subtropical, as well as humid tropical climates, although presumably this was not a large shift.

Climate(place) = humid tropics *(from 11)*

& Water-supply(place) = abundant & precipitation(place) = abundant

& Temp-range(place) = [hot, very hot]

& Season-description(place) = no winter, even rain

<==> Grain-grown(place) = [rice, corn, millet]

Climate(Florida) = subtropical, humid tropics

Water-supply(Florida) = moderate abundant

Precipitation(Florida) = moderate abundant

Season-description(Florida) = mild winter, long summer, even rain

Grain-grown(Florida) = corn *(no value given for Temp-range)*

(12) Climate(place) = [humid tropics, subtropical]

& Water-supply(place) = [abundant, moderate]

& Precipitation(place) = [abundant, moderate]

& Temp-range(place) = [hot, very hot]

& Season-description(place) = [even rain, long summer, [no, mild] winter]

<==> Grain-grown(place) = [rice (ϕ = .8) , corn (ϕ = .8) , millet (ϕ = .2)]

This implication was then be applied to Louisiana, as follows:

Climate(place) = [humid tropics, subtropical] *(from 12)*

& Water-supply(place) = [abundant, moderate]

& Precipitation(place) = [abundant, moderate]

& Temp-range(place)= [hot, very hot]

& Season-description(place) = [even rain, long summer, no or mild winter]

<==> Grain-grown(place) = [rice (ϕ = .8) , corn (ϕ = .8) , millet (ϕ = .2)]

Climate(Louisiana) = subtropical

Water-supply(Louisiana) = abundant

Precipitation(Louisiana) = abundant

Temp-range(Louisiana) = hot

Season-description(Florida) = mild winter, long summer, even rain

(13) Grain-grown(Louisiana) = rice, corn γ = moderate

= millet γ = low

Since Louisiana is not a perfect match to this complex implication, the certainty of the conclusion is reduced somewhat. For example, Louisiana's climate is subtropical, where most of the places that were considered in

forming the generalization were humid tropical places. Also, the season description that best matched Louisiana and predicted a grain was Florida, although it had somewhat lower precipitation.

2.5 Using Counter Evidence to Decrease Generalization Certainty

It is interesting to note that a counterexample to each of the generalizations just described was readily available in the data. In fact, Subject 1 had discussed this earlier in his session. The example was Peru, which was listed as having a mountainous terrain, a dry climate, only moderate water supply (irrigated), rainfall only during the summer, and growing both rice and corn. This prevented the subject from drawing conclusions about the terrain of the West Indies by reasoning backward from the grains grown there (rice, corn), but apparently the generalization in the forward direction was not affected. Here is a piece of the earlier interaction about the West Indies.

Protocol 7

S1: I can't infer anything about terrain here. One could say, well, it's like Java in growing rice and corn, maybe it has mountains and plains. [...] Since I believe that terrain affects grain growing. But this is a real weak inference. I don't like it, I mean I don't know. Maybe... but the rice and corn also is true here (Peru) and they had mountains. I mean you could look at the corn one.. see here (Florida) there are lowlands and plains. [...] So it looks as if corn can be grown in lowlands, but here (Peru) it took mountains. Corn and rice and you got mountains. So that tells you mountains and lowlands...I mean that is like two extremes. The evidence is contradictory as far as I am concerned.

This was a case where an attempted series of generalization refinements failed, because it ended with an attempt to generalize to a very dissimilar case. The effect of this is like attempting to combine two

inferences concluding in opposite directions, an inability to conclude anything.

Subject 2 made a similar attempt to refine an implication, leading to much the same conclusion, when reasoning about terrain in different places. His example also includes several different kinds of implication generalization/refinement rules that are part of the new generalization theory discussed in the next section. The subject first voices the implication that if a place grows corn it tends to be plains. Where this implication came from is not indicated at all in the protocol, but our conjecture is that he may have thought of places that grow corn, such as Illinois and Iowa, and noticed that they were on vast plains.

Protocol 8

E: What did you just figure out about the terrain of Angola or have you decided that you don't know?

S2: They grow corn. I would think normally that would tend to be plains. Check it out here. So in Florida they grow corn and it's planar. And in Java they grow rice and corn and its mountainous and lowlands. The lowlands could be plains I suppose. In Peru they grow corn and its mountainous, so that doesn't seem to be much of a help. So I guess I can't really conclude that on the whole it has plains. I'll skip it.

We formalize the first implication as having been formed by the example of the midwestern United States, as by:

Grain-grown(Iowa) = corn
Terrain(Iowa) = plains
Grain-grown(Illinois) = corn
Terrain(Illinois) = plains
Iowa, Illinois SPEC place

(14) Grain-grown(place) = corn <==> Terrain(place) = plains : γ = low

This is a simple generalization from common features to the effect that places with corn tend to be plains. Given the low certainty of that conjecture, the subject decided to check it out in the matrix to see if the places listed there which produce corn are plains. The first case he tried was Florida, and, indeed, its terrain had the value plains, so this increased the certainty of the implication. We call this an **Implication refinement for positive evidence** in the revised core theory described in Section 3:

Grain-grown(place) = corn <==> Terrain(place) = plains : γ = low
Grain-grown(Florida) = corn
Terrain(Florida) = plains
Florida SPEC place

(15) Grain-grown(place) = corn <==> Terrain(place) = plains : γ = moderate

Next, he considered the case of Java, where the terrain shown was mountains and lowlands. So Java fit the implication but it did not increase his certainty very much if at all. Finally, he considered the case of Peru (the last place where corn was listed). The terrain in Peru was mountains, which was clearly distinct from plains. This case *reduced his belief* in the implication below threshold, much as a DIS inference in the core theory would cancel a positive inference on the same question. The result of this was that he was unwilling to guess at the terrain of Angola on the basis of its growing corn. We call the inference about Peru an **Implication refinement for negative evidence** in the revised core theory:

Grain-grown(place) = corn <==> Terrain(place) = plains : γ = moderate
Grain-grown(Peru) = corn
Terrain(Peru) = mountains
mountains DIS plains
Peru SPEC place

(16) Grain-grown(place) = corn <==> Terrain(place) = plains : γ = very low

2.6 Conjunctive and Additive Combinations of Dependencies

As the examples above have indicated, this experiment has forced us to consider in more detail how dependencies and implications are affected by the presence of multiple contributing factors. In the core theory, when a variable (term) was dependent on several others, each dependency could be written separately (they could also be conjunctively combined in a single expression). In essence, this was the same as treating the independent terms disjunctively, since similarity-based inferences about the dependent term could be made based on similar values for any one of the independent terms alone. In many cases, subjects treat factors exactly this way. For example, Subject 1 reasoned about the water supply in Java (and other places) as being determined either by precipitation or rivers, independently.

Protocol 9

S1: Fresh water supply we don't know about ... It doesn't particularly have rivers, but it has a lot of abundant precipitation, so I assume that it has a good water supply... So I know it has abundant fresh water supply because it has abundant precipitation.

Another subject (2) took a different approach to combining these factors. When subject 2 tried to infer what the water supply for Florida was, he attempted to factor into his estimate for water supply the cumulative contributions of both rivers and rain. In the matrix, Florida was listed as having a sub-tropical climate, moderate to abundant precipitation, and mild winters with even rainfall throughout the year. However, no value was given for HAS-RIVER?.

Protocol 10

S2: Florida... The things that might affect fresh water supply are precipitation and whether there are rivers, according to this. We don't know if there are rivers for Florida, but there is abundant to moderate precipitation and that certainly is an important factor. So

I'd say on the basis of that partial evidence that its at least moderate water supply. I can't think of a way that I can figure out from this information whether there are rivers or not, so I guess I'd say moderate.

In contrast with the first subject, subject 2 appears to have treated precipitation as one of the two contributors to water supply. As a result, he discounted the value for precipitation somewhat when estimating water supply, because he didn't know if there were rivers in Florida. Apparently, he was unwilling to infer that there was abundant to moderate water supply given abundant to moderate precipitation without knowing that there were rivers in the place as well. This pattern repeats for the other countries where he made similar inferences, as shown below. In places where the matrix listed rivers as present, such as Saskatchewan, subject 2 inferred that the water supply was about the same as the precipitation level. In places like Iran and Java that had no rivers, his estimate of the water supply was consistently less than the precipitation level given for that place.

Protocol 11

S2: Saskatchewan is a dry climate and its got light precipitation, which sort of corresponds to that so I would say light ... light water supply. And I guess I would say light is greater than low.

S2: Iran has a semi-arid climate and light precipitation so I would say it probably has a less than moderate, low water supply.

S2: Java is a humid tropical climate with abundant precipitation and no rivers. But it probably still has... well, let's see. Rainfall evenly distributed over the year so I would say probably moderate to abundant water supply.

Much as he did when there was no value specified for rivers in Florida, this subject treated a NO value for HAS-RIVER as a reason to discount the precipitation figure in estimating water supply. For Iran, light

precipitation led to a low estimate for water supply (and he apparently treated low as less than light, based on what he said in protocol 10) because it was given as having no rivers. For Saskatchewan, he estimated a light, but greater than low, water supply because it was listed as having rivers and light precipitation. Similarly, Java, which had no rivers, was estimated to have only moderate to abundant water supply despite abundant precipitation.

In the examples from subject 2, above, it would appear that these are not simple conjuncts. Apparently, the subject has a qualitative model that is trying to combine the contributions of precipitation and rivers *additively*. This requires that the contribution of river water to the water supply be treated on a continuous scale. To formalize this, we introduce the implicit descriptor RIVER-WATER and use it instead of HAS-RIVER. The implications that relate these two descriptors we write as:

Has-river(place) = yes <==> River-water(place) ≥ moderate
Has-river(place) = no <==> River-water(place) = very low

We then write the dependency as:

River-water(place) + Precipitation(place) <--$^{+}$--> Water-supply(place)

Inferences using this kind of dependency require some mechanism for combining qualitative values that are scaled similarly. For example, the inference above about Java's water supply would look as follows, based on the protocol:

River-water(place) + Precipitation(place) <--$^{+}$--> Water-supply(place)
River-water(Java) = low
(17) Precipitation(Java) = abundant
Water-supply(Java) ≥ moderate

At this time, we are continuing to work out the details of how these kinds of inferences can be incorporated fully into the core theory.

3. Revisions to the Core Theory of Collins and Michalski

These protocol data, while consistent with the general framework of the core theory developed by Collins and Michalski (1989), led us to revise the theory to incorporate the way subjects induce new beliefs and the way subjects reason with "greater than" and "less than." At the same time we had begun to explore the space of all possible plausible inferences given the different kinds of expressions in the core theory. These two influences led to the revised theory presented in this section.

3.1 Reformulation of the Core Theory

Table 1 shows the types of expressions in the core theory as originally proposed. When we considered the kinds of generalizations that occurred in the protocols, it became clear that it was necessary to distinguish two kinds of mutual dependencies:

(1) unspecified mutual dependencies (D^u)

d_1 (a) <----> d_2 (a)

latitude (place) <----> temperature (place)

(2) specified mutual dependencies (D^s)

d_1 (a) $<--^{\pm}-->$ d_2 (a)

latitude (place) $<--^{-}-->$ temperature (place)

In the protocols subjects sometimes used unspecified mutual dependencies to guide their reasoning toward more specific relationships: either in the form of mutual implications or specified mutual dependencies. This distinction increases the number of expression types in the theory to five: relational statements (R), statements (S), mutual implications (I), unspecified mutual dependencies (D^u), and specified mutual dependencies (D^s). We still use D to represent cases where either type of dependency is possible.

Given these five types of expressions, it is possible to define a set of rewrite rules that take some expressions into other expressions. Table 7 shows these rewrite rules. The first entry shows that any GEN or SPEC relation can be rewritten as type and class statements. Since GEN and SPEC are inverse relations there are four rewrite rules shown. If one allows descriptors, such as "similar/to" and "dissimilar/to", there are also two rewrite rules for SIM and DIS.

The second entry in Table 7 shows how any statement can be rewritten as an implication. So for example the statement that cardinals are red can be rewritten in the form: if x is a cardinal, then its color is red. This rewrite rule makes clear that the frequency parameter ϕ in the core theory is the same as the conditional likelihood parameter α for the equivalent implication.

The third entry in Table 7 shows that any implication can be rewritten as an unspecified dependency. One can argue whether this is a rewrite rule or a generalization. If a dependency is interpreted as implying that the values of the variables (or terms) specified in the dependency are correlated over the entire range of the variables, then this rule is really a generalization from one part of the range to the entire range of the variables. But if a dependency is understood as stating simply that there is a correlation between the two variables, then a correlation over the limited range specified in the implication implies at least a weak overall correlation. Under this interpretation, the third entry is a rewrite rule.

The fourth entry shows that a specified dependency can be rewritten in terms of an implication where the values high, moderate, and low are specified appropriately. If there is more knowledge about the exact nature of the dependency (e.g. that temperature ~ 85^{o} when latitude ~ 0^{o}), then that can be incorporated into the rewritten implication.

At Michalski's suggestion, we attempted to construct all possible combinations of two or more premises leading to a conclusion given the five types of expressions. That is, we took each possible expression type

crossed with each possible expression type (2 premises), to see if it was possible to construct a plausible inference leading to each of the possible expression types as the conclusion [R, S, I, D^u, D^s] x [R, S, I, D^u, D^s) ----> [R, S, I, D^u, D^s]. Then we tried the same strategy with 3 premises. This exercise led to a reclassification of the types of plausible inferences in the core theory as shown in Table 8.

In Table 8 the substitutions have the form that any expression (E) can have an element in it replaced, if that element is related to another element by some relation (R). (Examples of each class are shown later in this section). All of the inferences about flowers shown in Table 2 were substitutions. The transitivity inferences combine pairs of implications (I) or dependencies (D) to yield a new implication or dependency. Derivations are inferences where, given that one side of an implication or dependency holds for a particular case, a person infers that the other side of the implication or dependency also holds for that case. These three classes were in the core theory of Collins and Michalski (1989) though the substitutions were called "transforms" and the transforms on implications and dependencies were treated as a separate class of inferences from statement transforms. This structure of the plausible inference space is more transparent than the one presented in the earlier paper.

The fourth class of plausible inferences, called generalizations, is new to the core theory. This class was prompted by the large number of inductions subjects were making in the experiment to form new hypotheses about the domain. The generalizations take two forms in the theory. One class, called simple generalizations, combine some number of statements and relations to form different types of expressions. The other class of generalizations, called refinements, start from an expression and use information about the world to further refine that expression. A number of examples of each kind of generalization will be shown below.

Table 8
Classes of Plausible Inferences

1. Substitutions

   E x R ——> E

2. Transitivities

   I/D x I/D —> I/D

3. Derivations

   I/D x S x R —> S

4. Generalizations (m and n indicate a variable number of expressions of a given form)

   Simple generalizations

   mS x nR —> E

   Refinement generalizations

   E x mS x nR —> E

Tables 9 through 15 show the different types of plausible inferences in the core theory together with examples. In the tables we have focussed on the critical premises to each plausible inference and have simplified the relational statements leaving out the context part of the statement. This is done for clarity of exposition: the proper form for most of these inferences is shown in Collins and Michalski (1989).

Table 9 includes the four types of substitution rules. The first type of rule substitutes one element (e.g. toads) for another element in a relation (e.g. frogs). This set of rules is new to the theory, but given rewrite rules that take relational statements into statements, it was implicitly part of the statement transforms of the original theory. The first rule in the set covers all the possible combinations except those involving GEN. We have disallowed DIS as the relation in the second premise because it leads to a conclusion with "Not," which we do not allow as a well formed expression in the theory. The last two rules in the set handle the special case of GEN, by treating GEN as the inverse of SPEC.

The second set of rules are simply the set of eight statement transforms shown in Table 2. Viewing them as substitutions where a_2 or r_2 are substituted for a_1 or r_1 in the original expression makes clear their relation to the other types of substitutions in the theory. The last two types of rules allow substitution of elements in implications and dependencies. For implications either the argument (a) or one of the referents (r) can be replaced; for dependencies only the argument can be replaced since there is no referent. These were called transforms on implications and dependencies in the original paper (Collins & Michalski, 1989).

Tables 10 and 11 show the two kinds of transitivity inferences and derivation rules in the theory. The transitivity rule on implications leads to the construction of a new implication linking two variables (e.g., grain and latitude) that were known to be linked to another variable (e.g., temperature). The transitivity rule on dependencies applies to either unspecified or specified dependencies. For specified dependencies, signs are combined following the rules for multiplication (same signs ---> plus,

opposite signs ---> minus). The two derivation rules, which are the most frequent plausible inferences in the protocols from the experiment, are shown in Table 11. The transitivity and derivation rules are the same as presented in Collins and Michalski (1989).

Table 9
Substitution Rules

$E \times R \longrightarrow E$

1) $R \times R \longrightarrow R$

$a_1\ REL_1\ a_2$
$a_3\ REL_2\ a_1$

$a_3\ REL_1\ a_2$

$a_1\ REL_1\ a_2$
$a_1\ GEN\ a_3$

$a_3\ REL_1\ b_2$

$a_1\ GEN\ a_2$
$a_3\ REL_2\ a_2$

$a_1\ GEN\ a_3$

frogs SPEC amphibian
toads SIM frogs
toads SPEC amphibian

Substitution in a Relational Statement

REL_1=SPEC,SIM,DIS
REL_2=SPEC,SIM

Table 9
Continued

2) S x R —> S Substitution in a Statement

$d(a_1)=r$
a_2 REL a_1 REL=SPEC,SIM,GEN

$d(a_2)=r$

$d(a_1)=r$
a_2 DIS a_1

$d(a_2)\neq r$

$d(a)=r_1$
r_2 REL r_1

$d(a)=r_2$

$d(a)=r_1$
r_2 DIS r_1

$d(a)\neq r_2$

means-of-locomotion(bird)={flying...}
bobolink SPEC bird
means-of-locomotion (bobolink)={flying...}

Table 9
Continued

3) I x R ---> I — Substitution in an Implication

$d_1(a)=r_1 <=> d_2(a)=r_2$
r_3 REL r_2 — REL=SPEC, GEN, SIM

$d_1(a)=r_1 <=> d_2(a)=r_3$

$d_1(a_1)=r_1 <=> d_2(a_1)=r_2$
a_2 REL a_1

$d_1(a_2)=r_1 <=> d_2(a_2)=r_2$

climate(place)=subtropical<=>fruit(place)={oranges...}
grapefruit SIM orange
climate(place)=subtropical<=>fruit(place)={grapefruit...}

4.) D x R ---> D — Substitution in a Dependency

$d_1(a_1) <---> d_2(a_1)$
a_2 REL a_1 — REL=SPEC, GEN, SIM

$d_1(a_2) <---> d_2(a_2)$

latitude(place)<--->temperature(place)
city SPEC place
latitude(city)<--->temperature(city)

Table 10
Transitivity Rules

1) I x I ---> I Transitivity on Implications

$d_1(a)=r_1 <=> d_2(a) = r_2$
$d_2(a)=r_2 <=> d_3(a) = r_3$

$d_1(a) = r_1 <=> d_3(a)=r_3$

grain(place) = rice <=> temperature(place) = high
temperature(place) = high <=> latitude(place)=low
grain(place) = rice <=> latitude(place) = low

2) D x D -----> D Transitivity on Dependencies

$d_1(a) <----> d_2(a)$
$d_2(a) <----> d_3(a)$

$d_1(a) <----->d_3(a)$

number of species(place) <----> temperature(place)
temperature(place) <---> latitude(place)
number of species(place) <----> latitude(place)

Table 11
Derivation Rules

1) $I \times S \times R \dashrightarrow S$ Derivation from an Implication

$d_1(a_1)=r_1 \Leftrightarrow d_2(a_1)=r_2$
$d_1(a_2) = r_1$
a_2 SPEC a_1

$d(a_2) = r_2$

grain(place)=rice<=>rainfall(place)=heavy
grain(Louisiana)=rice
Louisiana SPEC place
rainfall(Louisiana)=heavy

2) $D^S \times S \times R \longrightarrow S$ Derivation from a Dependency

$d_1(a_1) \leftarrow\!\!-^{+}\!\!\rightarrow d_2(a_1)$
$d_1(a_2) = \langle$high, medium, low$\rangle$
a_2 SPEC a_1

$d_2(a_2) = \langle$high, medium, low$\rangle$

$d_1(a_1) \leftarrow\!\!-^{-}\!\!\rightarrow d_2(a_1)$
$d_1(a_2) = \langle$high, medium, low$\rangle$
a_2 SPEC a_1

$d_2(a_2) = \langle$low, medium, high$\rangle$

temperature (place) <--->- latitude(place)
temperature (Amazon) = high
Amazon SPEC place
latitude (Amazon) = low

3.2 A Theory of Generalization

The generalization rules that are presented in Tables 12, 13, 14 and 15 are new to the core theory. Some of them were clearly seen in the protocols as subjects created new beliefs: for example, protocol 3 included a case where the subject formed a new SIM statement by looking at cases, and protocols 4 through 8 showed cases where subjects induced new implications. Given these cases, we have tried to construct a core theory of generalization that incorporates these cases into an overall structure that can generate the five kinds of expressions in the core theory. The attempt is to produce the minimal set of generalizations that in combination can account for the way the five expression types are formed by people.

Table 12 shows our conjecture for the minimal set of generalizations necessary to generate SPEC statements. The first rule simply allows the inference that if some instance (or subclass) has a particularly diagnostic feature of some class, then the instance is probably a member of the class. The multiplicity of the referent μ_r is our measure of diagnosticity: if the multiplicity is low, then not many other classes have that property. In the example we chose the S-curved neck as a diagnostic property of swans, but we could have chosen the entire body shape. If something has the shape of a swan, then it is probably a swan; though other evidence (as we shall see) may lead one to back off that hypothesis.

Table 12
Generalizations to form SPEC Statements

1) S x S ---> R — Initial Generalization

d(A) = r: γ_1, μ_r
d (a) = r: γ_2

---

a SPEC A: $\gamma = f(\mu_r, \gamma_1, \gamma_2)$

neck shape (swan) = S-curved
neck shape (x) = S-curved
x SPEC swan

2) R x 2S ---> R — Refining for Positive Evidence

a SPEC A: γ_1
d(A) = r: γ_2, μ_r
d(a) = r: γ_3

---

a SPEC A : $\gamma = \gamma_1 + f(\mu_r, \gamma_2, \gamma_3)$

x SPEC swan: γ_1
color (swan) = white
color (x) = white
x SPEC swan: $\gamma > \gamma_1$

Table 12
Continued

3) 2R x 2S ---> R Refining for Negative Evidence

a SPEC A: γ_1
d(A) = r_1: γ_2, μ_r
d(a) = r_2: γ_3
r_1 DIS r_2: γ_4

---

a SPEC A: $\gamma = \gamma_1$-f(μ_r, γ_1, γ_2, γ_3, γ_4)

x SPEC swan: γ_1
color (swan) = white
color (x) = black
black DIS white
x SPEC swan: $\gamma < \gamma_1$

4) R x 2S ---> R Rejecting for Negative Evidence

d(A) = r_1: γ_1, μ_r
d(a) = r_2: γ_2
r_1 DIS r_2: γ_3

---

class (a) ≠ A: γ = f (μ_r, γ_1, γ_2, γ_3)

color (swan) = white
color (x) = black
black DIS white
class (x) ≠ swan

The second rule in Table 12 shows how confirming evidence increases the certainty of the inference. If one thinks swans are white, and if an instance (or subclass) one thinks is a swan turns out to be white, then that increases certainty in the hypothesis that the instance is a swan. Again the increase in certainty depends on the multiplicity of the referent white. The third rule shows the parallel case of disconfirming evidence. If the object one is looking at is black, that would decrease the certainty of its being a swan for a person who believes swans are white. The fourth rule is an extension of the third rule; if a person has no reason to think an instance is a member of a particular class, then if it differs on any property of that class, it is evidence against it being in that class. Of course, in the world of everyday reasoning, negative evidence is never as certain as Popper (1969) might have us believe. Generalizations to form GEN statements are a simple variant on these rules for SPEC statements.

Table 13 shows the rules for forming SIM and DIS statements. Unlike the SPEC statements we have included the context (CX) part of the statement, because it is integral to the way that subjects seemed to be forming these statements in the protocols. The initial generalization to form a SIM statement occurred in several places in the protocols analyzed (e.g.. Protocol 4). Basically it involves identifying a descriptor (or variable) for which two cases have the same or similar referents and constructing the belief that the two cases are similar on that descriptor. So if Java and the West Indies both include humid tropics, then one can infer they are similar with respect to climate generally. The second rule parallels the SIM rule, for the DIS relation: If two cases differ on a particular descriptor, such as having short vs. long necks, one can form the statement that they are dissimilar in neck length. It is of course possible to have both DIS and SIM statements stored about the same cases (e.g. ducks and geese are similar with respect to feet and dissimilar with respect to necks).

The next two rules in Table 13 allow for refinement of SIM and DIS statements to incorporate two or more descriptors. For example, if Java and the West Indies are also similar in that both have abundant precipitation, then this can be added to the set of descriptors for which they are similar. The final rule allows for the generalization of the

descriptors on which two cases are related to a common superordinate descriptor. So, for example, one might induce that Java and the West Indies are similar with respect to all their climatological or geographical characteristics, based on their similarity with respect to climate and precipitation.

Table 13

Generalizations to Form SIM and DIS Statements

1) 2S x 3R ---> R — Initial Generalization to SIM

$d(a_1) = r_1$
$d(a_2) = r_2$
r_1 SIM r_2
a_1, a_2 SPEC A

---

a_1 SIM a_2 in CX (A,d)

climate (Java) = humid tropics
climate (West Indies) = {subtropical, humid tropics}
humid tropics SIM {subtropical, humid tropics}
Java, West Indies SPEC places
Java SIM West Indies in CX (places, climate)

2) 2S x 3R ---> R — Initial Generalization to DIS

$d(a_1)=r_1$
$d(a_2)=r_2$
r_1 DIS r_2
a_1, a_2 SPEC A

---

a_1 DIS a_2 in CX (A,d)

necklength (duck)=short
necklength (goose)=long
short DIS long
duck, goose SPEC birds
duck DIS goose in CX (birds, necklength)

Table 13
Continued

3) 4R x 2S ---> R Refining a SIM generalization

a_1 SIM a_2 in CX (A, d_1)
$d_2(a_1)=r_1$
$d_2(a_2)=r_2$
r, SIM r_2
a_1, a_2 SPEC A
a_1 SIM a_2 in CX (A, d_1 & d_2)

Java SIM West Indies in CX (places, climate)
precipitation (Java)=heavy
precipitation (West Indies)=abundant
heavy SIM abundant
Java, West Indies SPEC place
Java SIM West Indies in CX (places, climate & precipitation)

4) 4R x 2S ---> R Refining a DIS generalization

a_1 DIS a_2 in CX (A, d_1)
$d_2(a_1)=r_1$
$d_2(a_2)=r_2$
r_1 DIS r_2
a_1, a_2 SPEC A

---

a_1 DIS a_2 in CX (A, d_1 & d_2)

ducks DIS goose in CX (birds, necklength)
sound (ducks)=quack
sound (geese)=honk
quack DIS honk
ducks, geese SPEC birds
ducks DIS geese in CX (birds, necklength & sound)

Table 13
Continued

5) 2R x S ---> R

Descriptor Generalization

a_1 0 a_2 in CX(A, d_1 & d_2)
d_1, d_2 SPEC D

0 = any relation

---

a_2 0 a_2 in CX(A.D)

Java SIM West Indies in CX(places, climate & precipitation)
climate, precipitation SPEC climatological characteristics
Java SIM West Indies in CX(places, climatological characteristics)

Table 14 presents a set of four generalization rules for forming statements. The first rule is the simplest case of generalization from a subclass to a class; it is simply the argument generalization rule included in the statement substitutions in Table 9. The idea is that if you encounter a swan and it is white, one can infer that swans in general are white. The parameter v represents the number of swans encountered, be it one or a whole flock. Likewise, the second rule is the referent generalization rule included among the statement substitutions in Table 9.

The next two rules parallel the rules for refining evidence among the SPEC generalizations. The third rule is a refinement for negative evidence. If you think swans are white, and you encounter a black swan, this may lead to the idea that swans can be white or black. The frequencies one assigns to white swans and black swans depends on the number (v) of black and white swans encountered as is shown by the formulas. The fourth rule is a refinement for positive evidence, and it makes it possible to update the frequencies of different referent subclasses.

Table 14
Statement Generalizations

1) S x R —> S Argument Generalization

d(a) = r
A GEN a
d(A) = r

color(swan1)=white
swan GEN swan1
color (swan) = {white...}

2) S x R—>S Referent generalization

d(a) = {r_1...}
R GEN r_1

d(a) = {R...}

means of locomotion(people) = {walking...}
movement on foot GEN walking
means of locomotion (people) = movement on foot

Table 14
Continued

3) 2S x 2R —> S Refining for negative evidence

$d(A) = r_1: \phi_1, \nu_1$
$d(a) = r_2: \nu_2$
r_1 DIS r_2
a SPEC A

$$d(A) = \{r_1, r_2 \ldots\}: \phi_1' = \frac{\nu_1 \phi_1}{\nu_1 + \nu_2}, \phi_2' = \frac{\nu_2}{\nu_1 + \nu_2}$$

color(swan) =white: $\phi_1 = 1$
color(swan1) = black
white DIS black
swan1 SPEC swan
color (swan) = {white, black...}:$\phi_1' < 1$, $\phi_2' > 0$

4) 2S x R —> S Refining for positive evidence

$d(A) = \{r_1, r_2 \ldots\}: \phi_1, \phi_2, \nu_1$
$d(a) = r_1: \nu_2$
a SPEC A

$$d(A) = \{r_1, r_2 \ldots\}: \phi_1' = \frac{\phi_1 \nu_1 + \nu_2}{\nu_1 + \nu_2}, \phi_2' = \frac{\phi_2 \nu_1}{\nu_1 + \nu_2}$$

color (swan)={white, black...}: ϕ_1, ϕ_2
color (swan1) = white
swan1 SPEC swan
color (swan) = {white, black...}: $\phi_1' > \phi_1, \phi_2' < \phi_2$

Table 15 shows the set of generalizations we conjecture are sufficient to characterize the ways that humans create implications and dependencies. The first two rules in the set show how the common features or the contrasting features of two arguments can be used to construct implications and dependencies. These two rules are used by Socratic tutors in choosing cases for comparison by students (Collins & Stevens, 1982).

The first rule forms the hypothesis that if two arguments have two features (or referents) in common, then the two features are linked in some way. This is a rather uncertain inference. For example, if one believes that Japan is in Asia and produces rice, and that China is in Asia and produces rice, two possible conclusions follow from that: One is the implication that if a place is in Asia, it produces rice (and vice versa). The other is the dependency that the grains grown in a place depends on the continent of the place.

The second rule allows three different possible conclusions based on the fact that two arguments have contrasting features (or referents) with respect to two descriptors. For example, if you believe that South China grows rice and North China grows wheat, you might hypothesize three different implications or dependencies. One is that if a place is warm, it grows rice (and vice versa). Two is that if a place is cool, it grows wheat (and vice versa). Three is the dependency that the grain grown depends on the temperature of the place. These are somewhat more certain generalizations than the common-feature generalizations, but only marginally so. In general people might want to adduce more evidence for either in order for the certainty to achieve some threshold where they are willing to consider such hypotheses seriously.

The third rule was identified from the protocols, where subjects would instantiate dependencies in terms of the referents they identified for particular cases. In this way they would form new implications. In the example, if a person believes that temperature and grain are related, and that Saskatchewan which is cool produces wheat, the person might infer that in general places that are cool produce wheat (and vice versa).

Table 15
Implication and Dependency Generalizations

1) 4S x 2R ----> I/D^u Generalization based on common features.

$d_1(a_1)=r_1$
$d_2(a_1)=r_2$
$d_1(a_2)=r_1$
$d_2(a_2)=r_2$
a_1, a_2 SPEC A
$d_1(A)=r_1 <=> d_2(A)=r_2$
$d_1(A) <---> d_2(A)$

grain(Japan) = rice
continent(Japan) = Asia
grain(China) = rice
continent(China) = Asia
Japan, China SPEC place
grain(place) = rice <=> continent(place) = Asia
grain(place) <—> continent(place)

Table 15
Continued

2) 4S x 4R ---> I/I/D^{u} Generalization based on contrasting features

$d_1(a_1)=r_1$
$d_2(a_1)=r_2$
$d_1(a_2)=r_3$
$d_2(a_2)=r_4$
a_1, a_2 SPEC A
r_1 DIS r_3
r_2 DIS r_4

$d_1(A)=r_1 <=> d_2(A)=r_2$
$d_1(A)=r_3 <=> d_2(A)=r_4$
$d_1(A) <—> d_2(A)$

grain(South China) = rice
temperature(South China) = warm
grain(North China) = wheat
temperature(North China) = cool
South China, North China SPEC place
rice DIS wheat
warm DIS cool

---

grain(place)=rice<=>temperature(place)=warm
grain(place)=wheat<=>temperature(place)=cool
grain(place)<—>temperature(place)

Table 15
Continued

3) D^u x 2S x R ---> I Refining a dependency to form an implication

$d_1(A) <\!\!\longrightarrow d_2(A)$
$d_1(a) = r_1$
$d_2(a) = r_2$
a SPEC A
$d_1(A) = r_1 <\!\!=\!\!> d_2(A) = r_2$

grain(place)<—>temperature(place)
grain(Saskatchewan)=wheat
temperature(Saskatchewan)=cool
Saskatchewan SPEC place
grain(place)=wheat<=>temperature(place)=cool

4) I x 2S x 2R ---> I Refining an implication from negative evidence

$d_1(A) = r_1 <\!\!=\!\!> d_2(A) = r_2 : v_1, \phi_1$
$d_1(a) = r_1$
$d_2(a) = r_3 : v_2$
a SPEC A
r_3 DIS r_2
$d_1(A) = r_1 <\!\!=\!\!> d_2(A) = \{r_2, r_3\} : \phi_1' = \frac{v_1 \phi_1}{v_1 + v_2}, \phi_2' = \frac{v_2}{v_1 + v_2}$

grain(place)=wheat<=>temperature(place)=cool
grain(Italy)=wheat
temperature(Italy)=mild
mild DIS cool
Italy SPEC place
grain(place)=wheat<=>temperature(place)={cool,ϕ_1'=.5; mild, ϕ_2'=.5}

Table 15
Continued

5) I x 2S x R ---> I Refining an implication from positive evidence

$d_1(A) = r_1 <=> d_2(A)=\{r_2, r_3\}: \phi_1, \phi_2, v_1$
$d_1(a) = r_1$
$d_2(a) = r_2: v_2$
a SPEC A

$$d_1(A)= r_1<=>d_2(A)=\{r_2 , r_3\}: \phi_1 = \frac{\phi_1 v_1 + v_2}{v_1 + v_2}, \phi_2 = \frac{\phi_2 v_1}{v_1 + v_2}$$

grain(place)=wheat<=>temperature(place)={cool o=.5; mild ϕ=.5}
grain(North China)=wheat
temperature(North China)=cool
North China SPEC place
grain(place)=wheat<=>temperature(place)={cool o=.7; mild ϕ=.3}

6) I x 2R ---> I Refining an implication by referent combination

$d_1(A) = r_1 <=> d_2(A) = \{r_2, r_3 ...\}$
r_2, r_3 SPEC R

$d_1(A) = r_1 <=> d_2(A) = R$

grain(place)=rice<=>climate(place)={subtropical,tropical}
tropical, subtropical SPEC hot
grain(place)=rice<=>climate(place)=hot

The next two rules are used to refine implications. They parallel the earlier refinement rules for statements. Rule 4 refines an implication to incorporate negative evidence. For example, if one believes that places that produce wheat are cool (and vice versa) and one encounters Italy which has a rather mild climate, then one might infer that places that are cool or mild produce wheat, with frequencies representing the number of cases of each type one has encountered. Rule 5 similarly adjusts frequencies appropriately if one encounters positive evidence.

We think these generalization rules incorporate all the ways that subjects were forming new statements in the experiments. But we have written the rules to be as general as possible. People often make generalizations based on what would appear to be insufficient evidence, particularly in settings like our experiment, but they are constantly refining their generalizations, and often rejecting them (as in Protocol 6 with respect to millet) as too uncertain to take seriously. So the rule set we have developed will surely produce generalizations no one would believe if applied willy nilly. The way that people prevent making inappropriate generalizations is by using other knowledge inferentially to restrict the generalizations they make to beliefs consistent with what they know in general (see Collins & Michalski, 1989).

3.3 Reasoning with Ordered Variables and Inequalities

In analyzing the protocols from the experiment described in the last section, we were led to extend our core theory to deal with the issue of continuous or ordered variables and plausible reasoning with inequalities. The core theory of Collins and Michalski (1989) treated all referents (values for terms) as discrete values with no intrinsic relationships other than similarity and class membership. Clearly, this was a simplification. Variables like altitude, latitude, temperature, and even water-supply take referents that can be mapped onto numerical scales, given appropriate measurement techniques, and they may also be expressed qualitatively using terms like low, medium and high. Given a set of measurements for any one of these attributes, people quickly develop models of their normal ranges from observations, and develop approximate ranges on those scales that they can refer to using terms such as low, medium and high. These qualitative terms are treated as ordered, though they are not really mutually exclusive. Each qualitative value stands for a range or distribution of measured values, and those ranges may overlap to some degree. For instance, the ranges covered by low and medium might intersect in a small range called medium-low.

Ranges on ordered scales can be considered similar to the degree that they overlap. When one range overlaps the median or midpoint of another range and vice versa, then the two can be considered highly similar. For ranges and values that are not highly similar, we introduce the inequality relations $<$, $>$, $\leq$, and $\geq$. Although we have not done a careful study of how people interpret and compare these inexact ranges, for the sake of this paper, we will arbitrarily treat these relations as follows: The statement $d(a) > r$ means that the referent of d(a) is in the range from the midpoint of range r to the top of the scale that r is on. The statement $d(a) \geq r$ means that the referent of d(a) is in the range from the bottom of range r to the top of the scale that r is on. Similarly, given $d(a) = r_a$ and $d(b) = r_b$, $d(a) > d(b)$ is equivalent to $r_a > r_b$, and means that r_a is dissimilar from r_b, because the median of r_a is less than the bottom of r_b and the median of r_b is greater than the top of r_a.

Having introduced the notion of explicit orderings for referents that are ranges on ordered scales, the core theory as determined so far can be naturally extended to allow all plausible inferences that contain statements of the form d(a) = r to also allow the = to be replaced by any of <, >, ≤, and ≥. For example, in a SIM-based argument transform, we would rewrite the rule as:

$d(a_1) \sim r$
a_2 SIM a_1 in CX (A, D)
D(A) <-----> d(A)
$\underline{a_1, a_2}$ SPEC A
$d(a_2) \sim r$ where $\sim$ was one of =, < , >, ≤, or ≥.

In addition to this slight reformulation of the inference rules of the core theory, the introduction of inequalities yields some new generalization, transformation and derivation inferences involving these orderings. All of the new inferences involve specified dependencies (dependencies of the form $d_1(A)$<--$^{+}$-->$d_2(A)$ or $d_1(A)$<--$^{-}$-->$d_2(A)$), where increases or decreases in one referent value have corresponding effects on another. Collins and Michalski (1989) (p. 35) described rules for derivations from such dependencies between single terms, for the cases where the referent values were expressed as low, medium and high. Basically, for a positive dependency, a low value on $d_1(a)$ implies a low value on $d_2(a)$, medium goes to medium, and high to high. For negative dependencies, low goes to high, medium to medium, and high to low.

Table 16 shows the rules for *creating* inequalities with SIM-based transforms on dependencies and implications. These rules are much like the SIM-based referent transform rules described in Collins and Michalski (1989). In the SIM-based transforms of the core theory, two arguments or referents would be compared and found similar in a context (e.g., CX(A, D)), related to the left side of the dependency. In the new rules for generating inequalities from directed dependencies (rules 1 and 2 of Table 16), the two arguments a_1 and a_2 are related by an inequality instead of by similarity (as by $d_1(a_1) < d_1(a_2)$). To extend this rule to the case where d_1 was one of several factors that *together* determined d_3 (by either an

additive or conjunctive dependency), then all of the other factors d_2^* are required to be similar for a_1 and a_2 (or to differ in the same direction as they do for d_1). Rules 1 and 2 in Table 16 cover the cases where a_1 SIM a_2 in the context of these other descriptors d_2^*. The result is that values for $d_2(a)$ and $d_2(b)$ are ordered correspondingly for a positive dependency, and ordered in the reverse direction for a negative dependency. For example, using rule 2, if altitude and temperature are inversely correlated for places at similar latitudes, then low places (e.g., Miami) should be warmer than high places (e.g., Mexico City) at similar latitudes and vice versa. When some of these other features also vary, the inference becomes less certain. We are still developing a more complex model of how these further deviations from similarity affect the inference and its certainty.

Another kind of inference with inequalities corresponds to a derivation from an implication, as described in Collins and Michalski (1989). Rules 3 and 4 in Table 16 show these inference patterns. The difference from the normal derivations with implications is that there must also be a directed dependency to specify the direction of change between terms. Table 16 shows an example of the use of rule 3 that is based on the dependency between precipitation, rivers and water supply discussed in Section 2.7. If places with light precipitation and a river can be considered to have a moderate water supply (as Egypt and Italy were described in the protocol experiment), then one can conclude that a place like France with rivers and greater amounts of precipitation should have a greater overall water supply, because of the directed dependency bearing on water supply.

As mentioned above, both of these inference forms have the requirement that other relevant "contextual factors" are held constant. In the dependency-based transform rules (Table 16, rules 1 and 2), this is captured in the requirement that a_1 and a_2 are similar in the context of other terms affecting the target descriptor $d_3(A)$, (represented by

Table 16
Inequality-generating Inferences

$D \times D^S \times 2R \times 3S \rightarrow S$ Inequality Transforms with Directed Dependencies

(1) $d_1(A) \langle\text{--}^{+}\text{--}\rangle d_3(A)$
$d_1(A)$ & $d_2(A)^*$ $\langle\text{-----}\rangle$ $d_3(A)$
a_1, a_2 SPEC A
a_2 SIM a_1 in CX(A, $d_2(A)^*$)
$d_1(a_1) = r_1$
$d_1(a_2) \sim r_1$
$d_3(a_1) = r_3$

$d_3(a_2) \sim r_3$

(2) $d_1(A) \langle\text{--}^{-}\text{--}\rangle d_3(A)$
$d_1(A)$ & $d_2(A)^*$ $\langle\text{-----}\rangle$ $d_3(A)$
a_1, a_2 SPEC A
a_2 SIM a_1 in CX(A, $d_2(A)^*$)
$d_1(a_1) = r_1$
$d_1(a_2) \sim r_1$
$d_3(a_1) = r_3$

$d_3(a_2) \sim^{-} r_3$

Notes: $d_2(A)^*$ stands for all other terms that $d_3(A)$ depends on.
$\sim$ is one of $<$, $>$, $\leq$, or $\geq$, consistently within a rule, and $\sim^{-}$ is its inverse.

altitude(place) $\langle\text{--}^{-}\text{--}\rangle$ temperature(place) in CX(places, latitude)
altitude(place) & latitude(place) $\langle\text{-----}\rangle$ temperature(place)
Miami, Mexico City SPEC place
Miami SIM Mexico City in CX(places, latitude)
altitude(Mexico City) = high
altitude(Miami) < high
temperature(Mexico City) = moderate
temperature(Miami) > moderate

Table 16
Continued

$D^S \times I \times R \times 2S \rightarrow S$ Derivations from Implications with Directed Dependencies

(3) $d_1(A) \leftarrow\!\!-^{+}\!\!\rightarrow d_3(A)$
$d_1(A) = r_1$ & $d_2(A) = r_2^*$ <==> $d_3(A) = r_3$
a SPEC A
$d_2(a) = r_2$
$d_1(a) \sim r_1$

$d_3(a) \sim r_3$

(4) $d_1(A) \leftarrow\!\!-^{-}\!\!\rightarrow d_3(A)$
$d_1(A) = r_1$ & $d_2(A) = r_2^*$ <==> $d_3(A) = r_3$
a SPEC A
$d_2(a) = r_2$
$d_1(a) \sim r_1$

$d_3(a) \sim^{-} r_3$

Notes: $d_2(A) = r_2^*$ stands for all other terms in the implication.
$\sim$ is one of $<$, $>$, $\leq$, or $\geq$, consistently within a rule, and $\sim^{-}$ is its inverse.

precipitation(place) & has-river(place) $\leftarrow\!\!-^{+}\!\!\rightarrow$ water-supply(place)
precipitation(place) = very light & has-river(place) = yes <==>
water-supply(place) = moderate
France SPEC place
has-river(France) = yes
precipitation(France) > very light
water-supply(France) > moderate

$d_2(A)$*<-----> $d_3(A)$). In the rule for a plausible derivation from an inequality and an implication, this requirement is reflected in the need for the referents of all other terms on the left hand side of the implication (& $d_2(A)= r_{2*}$ ...) to be similar to the corresponding referent of $d_2(a)$ for the target example a, or ordered in the same direction. In the example presented with the table, the inequality derivation only works because France has rivers, as required by the left hand side of the implication.

Of course, there must also be ways of forming specified dependencies. The most straightforward of these looks much like the formation of a dependency by generalizing on contrasting features (see Table 14, part 2). This is shown in Table 17 for the simple case of comparing attributes of two exemplars. As an example of this type of generalization (Table 17, Rule 1) we show how one might derive the dependency that the latitude of a place is inversely correlated with its temperature. Comparing Alaska and Equador on these variables, we see that Alaska has a much higher latitude and a much lower average temperature than Equador. Generalizing on these facts gives the negative dependency. Much the same rules can be used to refine an unspecified dependency, essentially using the dependency to pick out the attributes that need to be compared. For instance, the generalization from Alaska and Equador would be made more certain if one already knew that there was a dependency between latitude and temperature, but the form of that relationship was unknown until the exact data was considered.

Another way to derive a directed dependency is by using two implications that address the same pair of descriptors. These generalization rules are much the same as rules 1 and 2, with pairs of statements rewritten as implications over the class A containing a_1 and a_2. An example of rule 3 is shown in Table 17. As in the previous example, the referents of the corresponding terms in the implications are placed in correspondence and their values compared. Since the direction of shift from tropical place to polar place is opposite on the two descriptors, a negative dependency is formed.

Table 17

Generalizing Based on Inequalities

4S x 2R --> D^S Generalizing to Specified Dependencies

(1) $d_1(a_1) = r_1$
$d_2(a_1) = r_2$
$d_1(a_2) = r_3$
$d_2(a_2) = r_4$
a_1, a_2 SPEC A
$r_1 \sim r_3$
$r_2 \sim r_4$

$d_1(A) <--^{+}--> d_2(A)$

(2) $d_1(a_1) = r_1$
$d_2(a_1) = r_2$
$d_1(a_2) = r_3$
$d_2(a_2) = r_4$
a_1, a_2 SPEC A
$r_1 \sim r_3$
$r_2 \sim^{-} r_4$

$d_1(A) <--^{-}--> d_2(A)$

$\sim$ is one of $<$, $>$, $\leq$, or $\geq$, consistently within a rule, and $\sim^{-}$ is its inverse.

latitude(Alaska) = high
temperature-range(Alaska) = cold
latitude(Equador) = low
temperature-range(Equador) = hot
Equador, Alaska SPEC place
hot > cold
low < high
latitude(place) <--$^{-}$--> temperature-range(place)

Table 17
Continued

$D \times 2I \times 2R \times 2S \rightarrow D^S$	Generalizing Implications to Directed Dependencies
(3) $d_1(A) \langle\text{----}\rangle d_2(A)$ $d_1(a_1) = r_1 \langle\text{==}\rangle d_2(a_1) = r_2$ $d_1(a_2) = r_3 \langle\text{==}\rangle d_2(a_2) = r_4$ $r_1 \sim r_3$ $r_2 \sim r_4$ $\underline{a_1, a_2 \text{SPEC } A}$ $d_1(A) \langle\text{--}^{+}\text{--}\rangle d_2(A)$	(4) $d_1(A) \langle\text{----}\rangle d_2(A)$ $d_1(a_1) = r_1 \langle\text{==}\rangle d_2(a_1) = r_2$ $d_1(a_2) = r_3 \langle\text{==}\rangle d_2(a_2) = r_4$ $r_1 \sim r_3$ $r_2 \sim^{-} r_4$ $\underline{a_1, a_2 \text{SPEC } A}$ $d_1(A) \langle\text{--}^{-}\text{--}\rangle d_2(A)$

latitude(place)$\Leftrightarrow$temperature-range(place)
latitude(tropical-place)=low$\Leftrightarrow$temperature-range(tropical-place)=hot
latitude(polar-region) = high$\Leftrightarrow$temperature-range(polar-region)=cold
low < high
hot > cold
tropical-place, polar-region SPEC place
latitude(place)$\langle\text{--}^{-}\text{--}\rangle$temperature-range(place)

The use of inequalities with qualitative values and other kinds of inexact or "fuzzy" categories has been studied by Zadeh (1965) and others using his theory. Our extension of the core theory to these kinds of statements and inferences was quite natural, and many of the implications of this extension were understood beforehand. Nonetheless, it raised a number of issues that are yet to be resolved, some of which are touched on in the next section. For example, there is a tradeoff between the precision and certainty of a referent value that pervades this kind of reasoning. We do not yet understand how and when people prefer a precise but uncertain answer as opposed to an imprecise but more certain one. The purpose for which the question was asked surely plays an important role here, but the mechanism by which that affects people's inference processes remains an open question.

4. Conclusion

In the revised theory we have addressed the issues raised by the experiment that we could find solutions for. There are a number of other issues apparent to us in the experiment and earlier protocols that we have not yet addressed. We think they are amenable to the kind of analysis we have been using, but the solutions were not as apparent or we did not have the time to pursue them. We enumerate them here so the reader can see what we think we have swept under the rug for the time being:

1. Combining variables on one side of a dependency or implication. In the experiment subjects frequently reasoned backwards or forwards over dependencies and implications where a number of variables (e.g. precipitation and rivers) affected a particular variable (e.g. water supply). Sometimes subjects treated the variables as if they were ORed together, and sometimes as if they were ANDed together, and sometimes as if they were additive. It is possible these reasoning patterns can be handled by a single combination rule with different α and β parameter values. Alternatively, it may be necessary to develop a slightly different set of plausible inference rules to handle each kind of combination. We simply have not resolved the issue to our satisfaction.

2. The tradeoff between range and certainty. Subjects appear to trade off certainty about a belief against the range of the referent. For example, one might be very certain that the average rainfall in Louisiana is "at least moderate," somewhat less certain that it is "heavy," still less certain that it is between 40 and 60 inches a year. In other words, for any continuous variable subjects can always increase their certainty in a belief by extending the range of the referent. Currently there is no way to incorporate such tradeoffs in the theory.

3. Merging of qualitative and quantitative reasoning. Sometimes subjects bring in various quantitative relationships to guide their qualitative reasoning (e.g. the Amazon jungle averages 85° or 1 mile in altitude affects temperature as much as 800 miles in latitude). There needs to be a smooth way to incorporate such quantitative information into the way humans reason plausibly.

4. Combining certainty parameters. Collins and Michalski (1989) carefully avoided specifying how people combine certainty parameters to arrive at an overall certainty in the conclusion. In this paper we did specify how the numeracy parameter v can logically be combined to derive frequency o. It should be possible to work out a normative theory for combining all the parameters specified in the theory, but we have not attempted to do so yet.

5. The extent parameter. Collins (1978) identified a parameter he called "extent" which was particularly prevalent in temporal and spatial inferences. It is necessary because people have a notion of how far rainstorms vs. parades vs. continents extend in space, and how long they extend in time. This notion is central to people's reasoning about space and time, but it also affects inferences in the core theory. For example, internal organs extend over a wider range of animals than horns or colors, so a person is more likely to infer that an animal has a gizzard because a similar animal has one, then one is to infer that an animal has a horn because a similar animal has one. We have not tried to incorporate this notion of extent into the core theory.

6. Finding relevant information in memory. The core theory of Collins and Michalski (1989) assumed that information is found by a marker passing search, and its impact on any question was evaluated by the plausible reasoning theory. There is a suggestion in the data from the experiment that each piece of information that is found redirects the search process in memory. We think therefore that it is possible to specify in more detail the

nature of the search to find relevant information to answer any question, but we have not yet worked out the details in this revision of the core theory.

7. Spatial, temporal, and meta inferences. As stated in the core theory of Collins and Michalski (1989) the protocols are full of plausible inferences based on spatial, temporal, and meta knowledge. We think an extension of the core theory to cover these inferences is possible, but it is a major enterprise that we still are not ready to tackle.

In summary we think the experimental data suggest that we are in the right ball park for constructing a general theory of human plausible reasoning. However, there is still much work to be done to accomplish this goal.

5. Acknowledgements

This research was supported by the Army Research Institute under Contract No. MDA903-85-C-0411.

We thank Ryszard Michalski for the many ideas he has contributed to our thinking about plausible reasoning and Judith Orasanu for her ideas and support in carrying out the research.

6. 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, 1, (pp. 185-188). Los Altos, CA: Morgan Kaufmann.

Burstein, M. H. & Collins, A. (1988). Modeling a theory of human plausible reasoning. In O'Shea, T. & Sgurev, V. (Eds.), Artificial intelligence III: Methodology, Systems, Applications (pp. 21-28). Amsterdam: North Holland.

Collins, A. (1978). 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. & Michalski, R. S. (1989). The logic of plausible reasoning: A core theory. Cognitive Science, 13, 1-49.

DeJong, G. F. (1981). Generalizations based on explanations. In Proceedings of the 7th International Joint Conference on Artificial Intelligence (pp. 67-69). Los Altos, CA: Morgan Kaufmann.

Laird, J. E., Rosenbloom, P. S., & Newell, A. (1986). Chunking in SOAR: The anatomy of a general learning mechanism. Machine Learning, 1, 11-46.

Michalski, R. S. (1987). How to learn imprecise concepts: A method for employing a two-tiered knowledge representation in learning. In Proceedings of the Fourth International Workshop on Machine Learning (pp. 50-58). Los Altos, CA: Morgan Kaufmann.

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.

Mitchell, T. M. (1983). Learning and problem solving. In Proceedings of the 8th International Joint Conference on Artificial Intelligence (pp. 1139-1151). Los Altos, CA: Morgan Kaufmann.

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 Kaufmann.

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