THE JOURNAL OF THE LEARNING SCIENCES, *1*(3 & 4), 319-359

Plausible Generalization: Extending a Model of Human Plausible Reasoning

Mark H. Burstein
Bolt Beranek and Newman Inc.
Cambridge, MA

Allan Collins
Bolt Beranek and Newman Inc.
and the
Institute for Learning Sciences
Northwestern University
Evanston, IL

Michelle Baker
Columbia University
New York

Transcripts of people answering questions or carrying on dialogues about everyday matters are filled with plausible inferences—inferences that are not certain, but that make sense. The same patterns of inferences occur in many different contexts. Often, in forming these inferences, people make generalizations that are equally uncertain but nevertheless are useful guides to reasoning. This article describes some important extensions to our earlier description of a core theory of plausible reasoning, based in large part on a new protocol study. The extensions are both data driven and theory driven. The primary focus here is on the inductive inference patterns people use to form plausible generalizations, that is, weakly held beliefs based on few examples but annotated with the same forms of certainty and similarity information that supported the inferential patterns described in our earlier work. We also provide examples of qualitative reasoning with inequalities and extend our formalism to cover that type of reasoning.

When looking at transcripts of people answering questions or carrying on dialogues about everyday matters, one notes that their comments are filled

Requests for reprints should be sent to Mark Burstein, Bolt Beranek and Newman Inc., 10 Moulton Street, Cambridge, MA 02238.

with plausible inferences—inferences that are not certain, but that make sense. It is striking that the same patterns of inferences occur in many different contexts. For a number of years, we have been characterizing the patterns of plausible inferences that occur in natural discourse and the various factors (certainly parameters) that make people more or less certain about the conclusions they are drawing (Baker, Burstein, & Collins, 1987; Burstein & Collins, 1988; Carbonell & Collins, 1973; Collins, 1978; Collins & Michalski, 1989; Collins, Warnock, Aiello, & Miller, 1975). In a recent article (Collins & Michalski, 1989), we characterized a core theory of plausible reasoning in terms of Michalski's (1987) variable-valued logic notation. This article introduces a learning theory based on plausible generalization using the formalism developed by Collins and Michalski (1989).

The work has been both data driven and theory driven. We began by characterizing the different patterns of inferences that occur in natural discourse. However, as we noticed relationships between different inference types and the parameters that affect certainty among the different inference types, we began identifying the overall structure of the plausible inference space. This was done by identifying different patterns from data and then generating new patterns that are variations of the patterns originally seen in the data to determine whether these new sets also produce plausible inferences. The overall goal, then, is to characterize a space of plausible inferences that derives from human data.

We can illustrate the responses we analyzed in terms of two protocols from the earlier study (Collins & Michalski, 1989). The first protocol comes from a teaching dialogue on South American geography:

Protocol 1

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

Tutor: I think it's more sheep country. It's like western Texas, so in some sense I guess it's cattle country.

At first, the tutor tentatively rejects the possibility of the Chaco as being cattle country because it is sheep country. This is called a *dissimilarity transform* in the theory; cattle country and sheep country are dissimilar enough that the tutor thinks cattle are unlikely. However, a *similarity transform* then leads to an affirmative conclusion, which partially counters the initial negative conclusion. The Chaco is similar to western Texas with respect to the variables that affect cattle raising (such as vegetation and climate), so it might be possible to raise cattle there. This protocol 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 *specializaton transform* in the theory. He thinks that the Andes Mountains are in most South American countries, so the mountains are likely to be in Uruguay. Two certainty parameters in the theory show up here: The higher the *frequency* of countries that have the Andes, and the more *typical* Uruguay is, the more certain the inference.

The extensions to the core theory of Collins and Michalski (1989) described in this study are based on a new set of protocols collected as subjects tried to reason about geographical attributes. This experiment provided lengthy examples of plausible reasoning from given and unknown information. The resulting protocols forced us to consider how people reason about quantities using "less than" and "greater than" and how they generalize to form new knowledge.

The second part of this article introduces a theory of generalization to accompany our plausible inference theory. Learning and generalizing go hand in hand with plausible inference for a variety of reasons. For instance, we observed in our protocol experiments that many of the inferences the subjects made were based on generalizations formed either while searching for an answer to a question or while considering a related question in the same session. Conversely, subjects often attempted to "verify," or increase, their certainty in their answers by looking for analogous examples with similar conclusions. Upon finding such examples, they automatically generalized the cases together rather than simply stating their conclusions with more conviction.

We also believed it was important to describe a generalization theory for our formalism to clarify our assumptions about the source of information required for plausible inference patterns. Because our formalism uses somewhat nonstandard representations that include frequency and likelihood information, we found it necessary to show, at least in principle, how that kind of information can be learned routinely.

COMPARISON WITH OTHER WORK ON UNCERTAIN REASONING

The theory we have developed has many similarities, at its lower levels, to several of the more popular formalisms for evidential reasoning: Bayesian inference (Pearl, 1988) and Dempster–Shafer theory (Dempster, 1967; Shafer, 1976). However, as Pearl (1990) pointed out, Bayesian methods require the specification of a complete probabilistic model that relates the set of hypotheses to the set of anticipated observations. Although this allows the model to be used to answer any probabilistic query covered by the model, it goes against the notion of plausible reasoning, where one is attempting to answer questions without having all of the pertinent information and, therefore, one is reasoning essentially by analogy. Dempster–Shafer theory, on the other hand, attempts to compute probabilities of necessity and provability instead of probabilities of truth. In principle, therefore, it is more like plausible reasoning. A major distinction between that formalism and the plausible reasoning model presented here is the introduction of different certainty parameters. By distinguishing and reasoning about dominance and typicality in class/subclass relationship and about the expected multiplicity of relationships and by combining these with context-sensitive measures of similarity and inference rule certainty, we are modeling the rich set of (sometimes conflicting) measures of conceptual relationships that appear to be used by people in evaluating the certainty of their inferences.

Another thread in this work that distinguishes it from these other evidential combination theories is the importance of the notion of dependencies. Dependencies are strongly related to Russell's (1989) notion of determinations. Both model a nonspecific relationship between concepts so that, by inferential combination with an example, one can reason by analogy to draw a plausible conclusion. This notion forms the basis of a large fraction of the plausible inferences we have observed and is the fundamental mechanism by which background knowledge of a general nature is used to narrow the focus and provide a proper context for a variety of plausible inferences. Dependencies are also critical to our proposed extension of the core theory addressing issues of generalization. Used in this vein, we show by protocol examples that dependencies are often used in place of more specific causal rules to guide generalization in a manner much like that of explanation-based generalization (DeJong & Mooney, 1986; Mitchell, Keller, & Kedar-Cabelli, 1986). Note, though, that because dependencies do not allow one to "prove" the generalized proposition in advance, dependency-guided generalization is a form of nonmonotonic inference.

One of the key elements of our model of plausible inference is it reliance on context-bounded measures of similarity in developing measures of inference certainty. There is an enormous body of literature on the subject of similarity (e.g., the collection edited by Vosniadou & Ortony, 1989), and we cannot hope to consider it all here. However, several points are worth making. Our approach to computing similarity draws on the prototype-based categorization model of Smith and Osherson (1984, 1989). By that treatment, similarity was essentially a function of the number of shared and nonshared attributes the two concepts, weighted by the salience of those attributes. In our model, the salience of individual attributes and relations is replaced by a context for each inference, usually based on the dependencies guiding the inference. (See the examples in the next section, and Table 4 for further detail.)

We also introduced the use of *multiplicity* as a means of capturing knowledge of the range of possible values for an attribute, to cover cases in which one has knowledge that values exist that cannot be enumerated. This model was implemented in Plausible Reasoning Simulation System (PRSS; Baker, Burstein, & Collins, 1987). Although not explicitly considered in our current model, the recent work of Cohen and Loiselle (1988) and Huhns and Stephens (1989) on decompositions of semantic relations should provide a mechanism for extending our theory to capture more fine-grained similarities by acknowledging the contributions of "similar" but non-identical relations.

Because of the general nature of our work in attempting to map out the "space" of plausible inferences, we have not explicitly addressed the processing issues related to control of plausible inference in great detail. However, we believe that our use of dependencies to direct both the choices of exemplars and the measurement of similarity in context provides a foundation for describing many of the qualitative, context-specific mechanisms involved in uncertain reasoning for particular problem-solving tasks, as discussed with respect to expert systems by Cohen (1987). We are in strong agreement with Cohen's point, as embodied in his system on Management of Uncertainty in Medicine (MUM), that people engaging in plausible reasoning seek both confirming and disconfirming evidence and will draw plausible conclusions in both lines of reasoning before combining evidence to reach a conclusion.

Our assumptions about the role of memory in the search for dependencies and analogs around which to form plausible inferences are perhaps most compatible with cognitive models of induction in support of problem solving like that illustrated by the Plausible Induction (PI) model described by Holland, Holyoak, Nisbett, and Thagard (1987). In particular, a directed form of spreading activation is assumed to control both the consideration of relevant inference rules (implications and dependencies) and the selection of useful analogs for purposes of induction and generalization. Clearly, however, the details of the PI model require significant extension to fully model the plausible inference and certainty combination rules described here and in Collins and Michalski (1989).

THE CORE THEORY OF COLLINS AND MICHALSKI

The four types of expressions in the core theory of Collins and Michalski (1989) 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: generalization (GEN), specialization (SPEC), similarity (SIM), and dissimilarity (DIS). Each relational statement specifies a context (CX) where the first variable is the domain over which typicality or similarity is computed, and the second variable is the descriptor(s) with respect to which typicality or similarity is 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 mutual implications and mutual dependencies. A mutual implication specifies how one statement (or compound statement) is related to another. The example states that warm temperature and heavy rainfall imply rice growing and vice versa. A mutual

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 CS (birds, biological characteristics)
duck SIM goose in CX (birds, habitat)
duck DIS goose in CX (birds, neck length)

Mutual implications (I)
$d_1(a) = r_1 \Longleftrightarrow d_2(a) = r_2$
temperature(place) = warm & rainfall(place) = heavy $\Longleftrightarrow$
grain(place) = rice

Mutual dependencies (D)
$d_1(a) \overset{+}{\longleftrightarrow} d_2(a)$
average temperature (place) $\overset{-}{\longleftrightarrow}$ latitude (place)

dependency relates two terms for example, 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 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 dissimilarity transform). These eight transforms are one of four classes of plausible inference in the core theory (the other classes are illustrated in Table 5).

One of the central concerns of the theory of Collins and Michalski (1989) is to specify how different parameters affect the certainty that people draw from their plausible inferences. Table 3 shows how different parameters affect the certainty of the eight plausible inferences in Table 2.

- Typicality (τ) affects GEN and SPEC transforms. The more typical England is of Europe, or Surrey is of England, with respect to climate

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
Effects of Different Parameters on Statement Transforms

Transforms (From Table 2)		τ	σ	α	ϕ	δ	μ_a	μ_r	Target Node
	(1) GEN	+	0	+	+	+	+	0	Europe
Argument-based	(2) SPEC	+	0	+	+	+	0	0	Surrey
	(3) SIM	0	+	+	+	0	+	0	Holland
	(4) DIS	0	–	+	–	0	–	0	Brazil
	(5) GEN	+	0	+	+	+	0	+	Tropical plants
Reference-based	(6) SPEC	+	0	+	+	+	0	0	Yellow roses
	(7) SIM	0	+	+	+	0	0	+	Peonies
	(8) DIS	0	–	+	–	0	0	–	Bougainvillea

Note. The certainty parameters are: τ = typicality; σ = similarity; ϕ = frequency; δ = dominance; α, β = conditional liklihood; μ = multiplicty (argument, referent); + = higher values of parameter increase the certainty of the inference; – = higher values of parameter decrease the certainty of the inference.

(or any variable that affects flower growing), the more certain is 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 is 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 climate affects flower growing, the more certain are 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 makes sense only 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 (δ) applies to GEN and SPEC inferences and reflects the degree the subset makes up a large part of the set. For example, because Surrey is 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. Because many countries presumably have daffodils and 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. Because countries usually have many different types of flowers, μ_r is high, and the referent-based inferences are 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 Collins and 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 in which the tutor inferred that cattle might be raised in the Chaco because that region is 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 is 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 derived only fairly low certainty (γ = low) from this inference that cattle might be raised in the Chaco.

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 = high
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 = high
mountains (Uruguay) = Andes: γ = moderate	

In general, when an inference involves a dependency, the relationships between terms (GEN, SPEC, SIM, DIF) are specified within a context, denoted by CX, that restricts the attributes compared with those related by the dependency governing the inference. For example, in the similarity transform example in Table 4, the dependency between vegetation and livestock of a region guides the comparison of Chaco and western Texas. Because the inference concludes a fact about the livestock of Chaco, the comparison is made on the types of vegetation in the two regions. The overall certainty of the conclusion will vary depending on the typicality or similarity between the related terms in that context, not simply with respect to overall typicality or similarity (which is written as "all characteristics").

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

There are three other classes of plausible inferences in the core theory developed by Collins and Michalski (1989), which are exemplified in Table 5. First, there are derivations from implications and dependencies. For example, if a person believes warm places with heavy rainfall produce rice and that the Amazon River basin is warm and has heavy rainfall, one might infer that rice is probably grown there. Second, there are 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 are transforms on implications and dependencies. For example, if one believes that places with a subtropical climate produce oranges and that citrus fruit is a GEN of oranges, then one might infer that places, in general, with subtropical climates produce citrus fruits. The different variants of these three classes of inferences are detailed in Collins and Michalski (1989).

This summarizes the core theory developed earlier. Subsequent to the development of the core theory, we conducted an experiment using a technique developed by Baker, Burstein, and Collins (1987), which is described in the section An Experiment on Human Plausible Reasoning. The findings from the experiment have led to a generalization theory that we described in the section Extensions To the Core Theory.

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

Derivations from implications and dependencies	
temperature(place) = warm & rainfall(place) = heavy $\Longleftrightarrow$ grain(place) = rice	$:\alpha_1, \alpha_2$ = moderate, γ_1, γ_2 = high
temperature(Amazon) = warm	$:\gamma$ = high, ϕ = moderate
rainfall(Amazon) = heavy	$:\gamma$ = high, ϕ = high
Amazon SPEC place	$:\gamma$ = high
———	———
grain(Amazon) = rice	$:\gamma$ = moderate
Transitivity inferences on implications and dependencies	
humidity (place) $\stackrel{+}{\longleftrightarrow}$ average temperature (place)	$:\alpha$ = moderate, β = moderate, γ = high
average temperature (place) $\stackrel{-}{\longleftrightarrow}$ latitude (place)	$:\alpha$ = moderate, β = moderate, γ = high
———	———
humidity (place) $\stackrel{-}{\longleftrightarrow}$ latitude (place)	$:\alpha$ = low, β = low, γ = high
Transforms on implications and dependencies	
climate (place) = subtropical $\Longleftrightarrow$ fruit (place) = {oranges . . .}	$:\alpha$ = moderate, β = moderate, γ = high
citrus fruit GEN oranges in CX (fruit, growing conditions)	$:\gamma$ = high, τ = high
growing conditions (fruit) $\longleftrightarrow$ place (fruit)	$:\alpha$ = high, β = high, γ = high
———	———
climate (place) = subtropical $\Longleftrightarrow$ fuit (place) = {citrus fruit . . .}	$:\alpha$ = moderate, β = moderate, γ = high

AN EXPERIMENT ON HUMAN PLAUSIBLE REASONING

As we argued in the earlier article (Collins & Michalski, 1989), problems arise in constructing a theory of plausible reasoning from considering people's answers to questions. Three problems discussed in the earlier article were: (1) it is a highly inferential, post hoc analysis, (2) the theory is likely to be underconstrained because any constraints operating on the invocation of inferences are not apparent in the protocols, and (3) it is possible that the answers are produced by some other inferential process and that the verbal answers are mere rationalizations. Therefore, we argued that to test the theory it is necessary to compare human reasoning to a computer simulation of the model (Baker, Burstein, & Collins, 1987; Burstein & Collins, 1988) over the same data base. To do this, we gave human subjects a partially specified matrix of geographical variables crossed by countries, shown in Table 6, that we had developed for the computer simulation. Then, we

TABLE 6
Experimental Matrix of Geographic Data

12 × 9	*Climate*	*Water Supply*	*Grain Grown*	*Has River?*	*Precipitation*	*Season Description*	*Soil Type*	*Temperature Range*	*Terrain*
Afghanistan	?	?	None	?	?	?	Brown, gray	Hot, very hot	Mountains
Angola	?	Moderate, abundant	Corn	Yes	Abundant	Summer rain	Dark brown, gray	Hot	?
Egypt	Dry climate	Moderate (irrigated)	Wheat	Yes	Very light	?	Gray	Very hot	Plains
Florida	Subtropical Humid tropical	?	Corn	?	Moderate, abundant	Mild winer, Long summer, Even rain	?	?	Lowlands, plains
Iran	Semi-arid, Mediterranean	?	?	No	Light	Winter rain	Gray	?	?
Italy	Mediterranean	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
Saskat-chewan	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	?

interviewed five different scientists (who were not geographers) as they attempted to fill in the missing cells in the matrix.

Subjects were 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 ½ hr to 1½ hr for different subjects. Each session was recorded on audio tape and transcribed.

Although the task for subjects was less natural than the teaching dialogue and question answering tasks used earlier, we think much of the flavor of natural plausible reasoning was captured by the task. There are two reasons we believe this is so: First, the same kinds of inference forms that we have discussed in earlier studies (Collins, 1978; Collins & Michalski, 1989) were predominant in the reasoning on this task. Second, the subjects talked quite naturally and at length about why they made their guesses, so there does not seem to be anything artificial about the discourse they produced. The task did have two aspects that differentiated it from protocols we previously collected. Because the matrix was in front of the subjects, they could consider more cases and variables than the one or two usually possible. Second, because the subjects were asked a number of questions with respect to the same set of variables, they accumulated more knowledge about the same topic area in this context than usual. So, these protocols had more of a flavor of scientific investigation to form generalizations than earlier protocols.

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 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.

A Sample Protocol

One example of an issue that arises in plausible reasoning 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, two variables directly affected water supply. The principal variable was a qualitative value (light, moderate, abundant) for the average amount of precipitation of the country. The second variable indicated whether there were or were not rivers in the country (yes or no). For Italy, the water supply was listed as being moderate, and the column labeled Has River? had the value Yes.

Protocol 3

> 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 that there were mountains. The Mediterranean climate led the subject to infer that Italy had limited precipitation, whereas 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 derivations from implications (see Table 5), although the second inference requires an extension to the theory that we come back to shortly.

In the second half of the response, the subject based his inferences on the evidence that the fresh water supply attributed to Italy was moderate and 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 each factor's contribution to overall water supply, a question for which the subject presumably had little direct knowledge. The other problem is the lack of information, even qualitative, on the amount of water available from rivers. The matrix provided only that there were 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 were directly dependent on precipitation, independently of the presence of rivers. In general, either precipitation or rivers can 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 $\Longleftrightarrow$ Precipitation(place) > *"normal"*:
γ = moderate, α = moderate
Terrain(Italy) = {mountains . . .}: γ = high, μ_a = high, ϕ = moderate

---

(1*) Precipitation(Italy) > *"normal"*: γ = moderate

Climate(place) = Mediterranean $\Longleftrightarrow$ Precipitation(place) = light:
γ = high, α = high
Climate(Italy) = Mediterranean: γ = high, μ_a = high, ϕ = high

---

(2) Precipitation(Italy) = light: γ = high

Precipitation(place) $\overset{+}{\longleftrightarrow}$ Water-supply(place)
:β = moderate, γ = high
Water-supply(Italy) = moderate: γ = high, μ_a = high, ϕ = high

---

(3) Precipitation(Italy) = moderate: γ = moderate

Precipitation(place) $\overset{+}{\longleftrightarrow}$ Water-supply(place)
:α = high, β = moderate, γ = high
Has-rivers(place) $\overset{+}{\longleftrightarrow}$ Water-supply(place)
:α = moderate, β = low-moderate, γ = high
Water-supply(Italy) = moderate: γ = high, μ_a = high, ϕ = high
Has-rivers(Italy) = yes: γ = high, μ_a = high

(4*) Precipitation(Italy) = moderate: γ = low *(Discount 3)*

Water-supply(Egypt = moderate: γ = high, μ_a = high, ϕ = moderate
Water-supply(Italy) = moderate: γ = high, μ_a = high, ϕ = high
Has-rivers(Egypt) = yes: γ = high, μ_a = high, ϕ = moderate
Has-rivers(Italy) = yes: γ high, μ_a = high, ϕ = high

(5*) Italy SIM Egypt in CX(countries, Has-rivers & Water-supply):
γ = high, σ = high

Precipitation(place) OR Has-rivers(place) $\overset{+}{\longleftrightarrow}$ Water-supply(place)
:α_1, α_2 = high, β_1, β_2 = moderate, γ = high

Italy SIM Egypt in CX(countries, Has-rivers & Water-supply)
$: \gamma$ = high, σ = high *(from 5)*
Precipitation(Egypt) = very light: γ = high, μ_a = high, ϕ = high

---

(6) Precipitation(Italy) = very light: γ = moderate

Precipitation(Italy) > *"normal"*: γ = moderate *(from 1)*
Precipitation(Italy) = light: γ = high *(from 2)*
Precipitation(Italy) = very light: γ = moderate *(from 6)*

(7*) Precipitation(Italy) = light: γ = moderate-high

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 general issue of continuous variables and inequalities will be discussed later.

The first inference (1*) also raises an issue for the theory of Collins and Michalski (1989) that is implicit in the use of default values in the frame theory of Minsky (1975). 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 variables other than mountains, mountains tends to make the rainfall higher. So, if Mediterranean climates have light rainfall, the mountains make the rainfall greater than light. "Normal" is a "dummy value" for the precipitation variable used to carry forward the reasoning. This dummy value is filled in by Inferences 2 and 6, and the average value computed from those inferences is adjusted upward in Inference 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 (Pearl, 1987). The pattern occurs when there are several variables that independently influence a dependent variable. This can be written either as:

$$d_1(a) \text{ OR } d_2(a) \longleftrightarrow d_3(a)$$

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

$$d_1(a) \longleftrightarrow d_3(a) \text{ and } d_2(a) \longleftrightarrow d_3(a)$$

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

$$\frac{\begin{array}{ll} d_1(a) \longleftrightarrow d_3(a) & \\ d_3(a) = r & \text{for } r \in \{\text{high, medium, low}\} \end{array}}{d_1(a) = r}$$

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, the discovery that Italy has rivers and that this accounts for Italy's moderate water supply (by analogy to Egypt), thus, decreases the certainty of Inference 3 that 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). The set is 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 occurs in Inference 6 where the subject makes a generalization about how Italy and Egypt are similar. This rule, called an initial generalization to SIM, is part of the new generalization theory and appears in Table 8 as Rule 1.

Examples of Plausible Generalization

The next portion of the protocol of Subject 1 illustrates a new component of the theory, 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, in which Subject 1 reasoned about Egypt and Italy, and this protocol 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, refining a dependency to form an implication (Table 10, Rule 3), described in the section Extensions to the Core Theory. Generalizations like Inference 8, where an existing dependency is combined with specific examples 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 this class of generalizations *refinements*.

Reasoning With Inequalities

Once the generalization just described has been formed, the protocol given above shows how it is 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's and Italy's, means that it should have greater precipitation.

Has-river(place) & Precipitation(place) $\overset{+}{\longleftrightarrow}$ Water-supply(place)
Has-river(place) = yes & Precipitation(place) = light
$\Longleftrightarrow$ Water-supply(place) = moderate
Has-river(Louisiana) = yes
Water-supply(Louisiana) = abundant (or $>$ moderate)

---

(9) Precipitation(Louisiana) $>$ light (or $\geq$ moderate)

In Protocol 4, the subject reached the conclusion that because Louisiana had an abundant water supply, it has "at least moderate" precipitation. In the formalization of this inference, we describe the pattern as predicting simply that it is "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 class of inferences that was 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,* which is a dependency labeled with a + or − to indicate that an increase or decrease in the values for a term on one side has a corresponding positive or negative effect on the other. These inferences are formulated in the section Reasoning With Ordered Variables and Inequalities.

Reasoning Using Multiple Dependent Factors

When considering which cereal grains can be grown in different places, subjects were faced with a situation where all of the descriptors in the matrix are potentially relevant to some degree. Water supply, climate, soil type, and terrain all directly influence what grains are grown, and the other variables in the matrix contribute to those four.

When subjects made inferences where a number of contributing factors were known, the pattern that emerged was quite consistent. The subjects first formed generalizations and then used those generalizations to answer the questions. These generalization inferences occurred quite frequently and did not always help 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. Protocols 4 and 5 are clear examples of this. The second path is the formation of a weak generalization based on a single example and a dependency and then a refinement of that generalization as new confirming examples are encountered.

In the generalization theory presented in the section Extensions to the Core Theory, we treat these two paths as variants of the same general induction strategy because we wanted to produce a universal set of generalization rules. So, we assumed that the generalizations in Protocols 4 and 5 from two cases can be viewed as two-step inferences: refining a dependency to form an implication (Table 10, Rule 3) based on one case and refining an implication from positive evidence (Table 10, Rule 5) based on the other case. This eliminates the need of separate rules for one- and two-case generalizations.

In Protocol 5, Subject 1 searched for a way to answer questions on the terrain of the West Indies.

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 that the grains grown are rice and corn. We formalize the latter inference in the next paragraph. Because it is common knowledge that humid, tropical places are hot and wet, our formalization of this inference includes the factors of temperature, precipitation, and water supply as part of the generalization.

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)
$\longleftrightarrow$ Grain-grown(place)
Climate(Java) = humid tropics
Precipitation(Java) = abundant
Temp-range(Java) = hot
Grain-grown(Java) = rice, corn

---

(10) Climate(place) = humid tropics
& Precipitation(place) = abundant
& Temp-range(place) = hot
$\Longleftrightarrow$ Grain-grown(place) = rice, corn: γ_1

Climate(place) = humid tropics
& Precipitation(place) = abundant & Temp-range(place) = hot
$\Longleftrightarrow$ Grain-grown(place) = rice, corn: γ_1 *(from 10)*
Climate(West Indies) = humid tropics
Precipitation(West Indies) = abundant
Temp-range(West Indies) = hot

Grain-grown(West Indies) = rice, corn

---

(11) Climate(place) = humid tropics
& Precipitation(place) = abundant
& Temp-range(place) = hot
$\Longleftrightarrow$ Grain-grown(place) = rice, corn: $\gamma_2 > \gamma_1$

Using Counter Evidence to Decrease Generalization Certainty

Subject 2 similarly attempted to refine an implication and concluded that he did not have enough evidence when reasoning about terrain in different places. His example also includes several different 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. The source of this implication has no basis at all in the protocol, but our conjecture is that he applied previous knowledge of the geographical feature of places that grow corn, such as Illinois and Iowa, which are located on vast plains.

Protocol 6

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 it's mountainous and lowlands. The lowlands could be plains I suppose. In Peru they grow corn and it's 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 examples from the midwestern United States, as follows:

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

---

(11) Grain-grown(place) = corn Terrain(place) = plains: γ = low

This is a simple *generalization based on common features* (Table 10, Rule 1) to the effect that places with corn tend to be plains. Given the low certainty of that conjecture, the subject decided to determine from the matrix whether places listed there that 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 *refining an implication from positive evidence* in the system of generalization patterns described in the next section (Table 10, Rule 5):

Grain-grown(place) = corn $\Longleftrightarrow$ Terrain(place) = plains: γ = low
Grain-grown(Florida) = corn
Terrain(Florida) = plains
Florida SPEC place

---

(12) Grain-grown(place) = corn $\Longleftrightarrow$ Terrain(place) = plains: γ = moderate

Next, he considered the case of Java, and its terrain had the values of mountains and lowlands. So, Java fit the implication, but increased his certainty only very little if at all. Last, he considered the case of Peru (the final place where corn was listed). The terrain in Peru had the value of mountains, which is clearly distinct from the value of plains. We call the inference about Peru *refining an implication from negative evidence* in the generalization theory (Table 10, Rule 4):

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

---

(13) Grain-grown(place) = corn Terrain(place) = plains: γ = very low

This case reduced his belief in the implication below threshold, much as a DIS inference in the core theory cancels a positive inference on the same question. The result is that he was unwilling to guess at the terrain of Angola on the basis of its growing corn.

EXTENSIONS TO THE CORE THEORY

These protocol data, although consistent with the general framework of the core theory developed by Collins and Michalski (1989), led us to extend the theory to incorporate the way subjects induce new beliefs and reason with "greater than" and "less than."

A Theory of Plausible Generalization

The generalization rules that are presented in Tables 7, 8, 9, and 10 are new to the core theory. Some were clearly used in the protocols as subjects created new beliefs. For example, Protocol 3 showed the subject forming a new SIM statement by combining evidence from other sources, and Protocols 4 through 6 showed subjects inducing new implications. Given these cases, we constructed a core theory of generalization that incorporates these cases into an overall structure that generates the four kinds of expressions in the core theory. The attempt is to produce the minimal set of generalizations that in combination accounts for these expression types formed by people.

Table 7 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 reference 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.

The second rule in Table 7 shows how confirming evidence increases the certainty of the inference. If one thinks swans are white and if an instance (or subclass) believed to be a swan is 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 is black, that decreases the certainty of it being a swan for a person who believes swans are white. Generalizations to form GEN statements are a simple variant on these rules for SPEC statements.

Table 8 shows the rules for forming SIM and DIS statements. The initial generalization 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 generally similar with respect to climate. The second rule parallels the SIM rule for the DIS relation: If two cases differ on a particular descriptor, such as having short versus long necks, one can form the statement that they are dissimilar in neck length. Of course, it is possible to have both DIS and SIM statements stored for 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 8 allow for refinement of SIM and DIS

TABLE 7
Generalizations for Form SEPC Statements

(1) Initial generalization to SPEC

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

neck shape (swan) = S-curved	$:\gamma$ = high, μ_r = low
neck shape (x) = S-curved	$:\gamma$ = high
x SPEC swan	$:\gamma$ = low

(2) Refining a SPEC generalization for positive evidence

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

x SPEC swan	$:\gamma_1$
color (swan) = white	$:\gamma$ = high, μ_r = low
color (x) = white	$:\gamma$ = high
x SPEC swan	$:\gamma > \gamma_1$

(3) Refining a SPEC generalization for negative evidence

a SPEC A	$:\gamma_1$
d(A) = r_1	$:\gamma_2, \mu_r$
d(a) = r_2	$:\gamma_3$
r_1 DIS r_2	$:\gamma_4, \sigma$
a SPEC A	$:\gamma = \gamma_1 - f(\mu_r, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \sigma)$

x SPEC swan	$:\gamma_1$
color (swan) = white	$:\gamma$ = high, μ_r = low
color (x) = black	$:\gamma$ = high
black DIS white	$:\gamma$ = high, σ = low
x SPEC swan	$:\gamma < \gamma_1$

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 8

Generalizations to Form SIM and DIS Statements

(1) Initial generalization to SIM

$d(a_1) = r_1 \quad :\gamma_1, \mu_{r1}$

$d(a_2) = r_2 \quad :\gamma_2, \mu_{r2}$

r_1 SIM $r_2 \quad :\gamma_3, \sigma_1$

a_1, a_2 SPEC A $\quad :\tau_1, \tau_2, \gamma_4, \gamma_5$

---

a_1 SIM a_2 in CX (A,d)

$:\sigma = f_\sigma(\gamma_1, \mu_{r1}, \gamma_2, \mu_{r2}, \gamma_3, \sigma_1, \tau_1, \tau_2, \gamma_4, \gamma_5)\ \gamma = f_\gamma(\gamma_i)$

climate (Java) = humid tropics $\quad :\gamma$ = high, μ_r = low

climate (West Indies) = subtropical $\quad :\gamma$ = high, μ_r = low

humid tropics SIM subtropical $\quad :\gamma$ = high, σ = moderate

Java, West Indies SPEC places $\quad :\gamma_1, \gamma_2$ = high, τ_1, τ_2 = high

---

Java SIM West Indies in CX (places, climate)

γ = high, σ = moderate

(2) Initial generalization to DIS

$d(a_1) = r_1 \quad :\gamma_1, \mu_{r1}$

$d(a_2) = r_2 \quad :\gamma_2, \mu_{r2}$

r_1 DIS $r_2 \quad :\gamma_3, \sigma_1$

a_1, a_2 SPEC A $\quad :\tau_1, \tau_2, \gamma_4, \gamma_5$

---

a_1 SIM a_2 in CX (A,d)

$:\sigma = f_\sigma(\gamma_1, \mu_{r1}, \gamma_2, \mu_{r2}, \gamma_3, \sigma_1, \tau_1, \tau_2, \gamma_4, \gamma_5)\ \gamma = f_\gamma(\gamma_i)$

necklength (duck) = short $\quad :\gamma$ = high, μ_r = low

necklength (goose) = long $\quad :\gamma$ = high, μ_r = low

short DIS long in CX (necks, length) $\quad :\sigma$ = low, γ = high

duck, goose SPEC birds $\quad :\gamma_1, \gamma_2$ = high, τ_1, τ_2 = high

---

duck DIS goose in CX (birds, necklength) $:\sigma$ = low, γ = high

(3) Refining a SIM generalization

a_1 SIM a_2 in CX (A, d_1) $\quad :\sigma_1, \gamma_1$

$d_2(a_1) = r_1 \quad :\gamma_2, \mu_{r1}$

$d_2(a_2) = r_2 \quad :\gamma_3, \mu_{r2}$

r SIM $r_2 \quad :\gamma_4, \sigma_2$

a_1, a_2 SPEC A $\quad :\tau_1, \tau_2, \gamma_5, \gamma_6$

---

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

$:\sigma = f_\sigma(\sigma_1, \gamma_1, \gamma_2, \mu_{r1}, \gamma_3, \mu_{r2}, \gamma_4, \sigma_2, \tau_1, \tau_2, \gamma_5, \gamma_6)\ \gamma = f_\gamma(\gamma_i)$

Java SIM West Indies in CX (places, climate)

$:\gamma$ = high, σ = moderate

precipitation (Java) = heavy $\quad :\gamma$ = high, μ_r = low

precipitation (West Indies) = abundant $\quad :\gamma$ = high, μ_r = low

(continued)

TABLE 8 *(Continued)*

heavy SIM abundant in CX (precipitation, amount)
$:\gamma$ = high, σ = high
Java, West Indies SPEC places $:\gamma_1, \gamma_2$ = high, τ_1, τ_2 = high

---

Java SIM West Indies in CX (places, climate & precipitation)
$:\gamma$ = high, σ = moderate

(4) Refining a DIS generalization

a_1 DIS a_2 in CX (A, d_1)	$:\sigma_1, \gamma_1$
$d_2(a_1) = r_1$	$:\gamma_2, \mu_{r1}$
$d_2(a_2) = r_2$	$:\gamma_3, \mu_{r2}$
r_1 DIS r_2	$:\gamma_4, \sigma_2$
a_1, a_2 SPEC A	$:\tau_1, \tau_2, \gamma_5, \gamma_6$

---

a_1 DIS a_2 in CX (A, d_1 & d_2)
$:\sigma = f_\sigma(\sigma_1, \gamma_1, \gamma_2, \mu_{r1}, \gamma_3, \mu_{r2}, \gamma_4, \sigma_2, \tau_1, \tau_2, \gamma_5, \gamma_6)\ \gamma = f_\gamma(\gamma_i)$

ducks DIS goose in CX (birds, necklength): σ = low, γ = high
sound (ducks) = quack: γ = high, μ_r = low
sound (geese) = honk: γ = high, μ_r = how
quack DIS honk in CX (sounds, quality) $:\sigma$ = low, γ = high
ducks, geese SPEC birds: γ_1, γ_2 = high, τ_1, τ_2 = high

---

ducks DIS geese in CX (birds, necklength & sound): σ = low, γ = high

(5) Descriptor generalization

a_1 REL a_2 in CX(A, d_1 & d_2)	$:\gamma_1, \sigma, \tau_1$
d_1, d_2 SPEC D	$:\gamma_2, \gamma_3, \tau_2, \tau_3,$

---

a_2REL a_2 in CX(A,D) $:\gamma = f(\gamma_i, \sigma, \tau_1, \tau_2, \tau_3)$
where REL is any of (SIM, DIS, GEN, or SPEC)

Java SIM West Indies in CX(places, climate, & precipitation)
$:\gamma$ = high, σ = moderate
climate, precipitation SPEC climatological characteristics
$:\gamma_1, \gamma_2$ = high, τ_1, τ_2 = high

Java SIM West Indies in CX (places, climatological characteristics)
$:\gamma$ = high, σ = moderate

Table 9 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. The idea is that if one encounters a swam and it is white, one can infer that swans in general are white. The parameter ν 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 2.

The next two rules parallel the rules for refining evidence among the

TABLE 9
Statement Generalizations

(1) Argument generalization[a]

$d(a) = r$ $\quad :\gamma_1, \phi, \mu_a$
A GEN a in CX (A, D(A)) $\quad :\gamma_2, \tau, \delta$
$\{D(A) \longleftrightarrow d(A)$ $\quad :\alpha, \gamma_3\}$

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

color(swan1) = white $\quad :\gamma$ = high, μ_a = high
swan GEN swan1 in CX (swans, all characteristics)
$:\gamma$ = high, τ = high, δ = low

---

color (swan) = white $\quad :\gamma$ = moderate

(2) Referent generalization

$d(a) = \{r_1 \ldots\}$ $\quad :\gamma, \phi, \mu_r$
R GEN r_1 in CX (d, D(d)) $\quad :\gamma_2, \tau, \delta$
$\{D(d) \longleftrightarrow A(d)$ $\quad :\alpha, \gamma_3\}$
a SPEC A $:\gamma_4$

$d(a) = \{R \ldots\} :\gamma = \gamma_1 - f(\gamma_1, \mu_a, \gamma_2, \tau, \delta, \gamma_4)$

agricultural product (Honduras) = {bananas . . .}
$:\gamma$ = high, μ_r = moderate, ϕ = moderate
Tropical fruits GEN bananas $\quad :\gamma$ = high, τ = high, δ = low
Climate (agricultural products) $\longleftrightarrow$
Place (agricultrual products) $\quad :\gamma$ = high, α = high
Honduras SPEC place $\quad :\gamma$ = high

---

agricultrual product (Honduras) = {tropical fruits . . .}
$:\gamma$ = high

(3) Refining for negative evidence

$d(A) = r_1$ $\quad :\phi_1, \nu_1, \gamma_1$
$d(a) = r_2$ $\quad :\phi_2, \nu_2, \gamma_2$
r_1 DIS r_2 $\quad :\gamma_3, \sigma$
a SPEC A $\quad :\gamma_4, \tau, \delta$

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

$$\gamma = f(\gamma_1, \gamma_2, \gamma_3, \gamma_4, \sigma, \tau, \delta)$$

color(swan) = white $\quad :\phi_1 = 1$
color(swan1) = black $\quad :\gamma$ = high
white DIS black $\quad :\gamma$ = high, σ = low
swan1 SPEC swan $\quad :\gamma$ = high, τ = low

---

color (swan) = {white, black . . .} $\quad :\phi_1' < 1, \phi_2' > 0, \gamma$ = high

(continued)

TABLE 9 *(Continued)*

(4) Refining for positive evidence	
d(A) = {r_1, r_2 . . .}	:ϕ_1, ϕ_2, ν_1, γ_1
d(a) = r_1	:γ_2, ν_2
a SPEC A	:γ_3, τ, δ
d(A) = {r_1, r_2 . . .}	:$\phi_1' = \frac{\phi_1\nu_1 + \nu_2}{\nu_1 + \nu_2}$, $\phi_2' = \frac{\phi_2\nu_1}{\nu_1 + \nu_2}$,
	$\gamma = f(\gamma_1, \gamma_2, \gamma_3, \sigma, \tau, \delta)$
color (swan) = {white, black . . .}	:ϕ_1, ϕ_2
color (swan1) = white	:γ = high
swan1 SPEC swan	:γ = high, τ = high
color (swan) = {white, black . . .}	:$\phi_1' > \phi_1$, $\phi_2' < \phi_2$, γ = high

[a]From Table 7 in Collins and Michalski (1989).

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 (ν) of black and white swans encountered as is shown by the formulas. The fourth rule is a refinement for positive evidence, and it makes possible the updating of frequencies of different referent subclasses.

Table 10 shows that the 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 an uncertain inference. For example, of one believes that Japan is in Asia and produces rice and that China is in Asia and produces rice, two possible conclusions follow: One is the implication that if a place is in Asia, it produces rice (and vice versa). The other is the dependency that the types of grain grown in a place depend on which continent the place is located.

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 one believes that South China grows rice and that North China grows wheat, one 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

TABLE 10

Implication and Dependency Generalizations

(1) Generalization based on common features

$d_1(a_1) = r_1$ $\quad :\gamma_1, \mu_{r1}, \mu_{a1}, \phi_1$

$d_2(a_1) = r_2$ $\quad :\gamma_3, \mu_{r2}, \mu_{a2}, \phi_2$

$d_1(a_2) = r_1$ $\quad :\gamma_3, \mu_{r3}, \mu_{a3}, \phi_3$

$d_2(a_2) = r_2$ $\quad :\gamma_4, \mu_{r4}, \mu_{a4}, \phi_4$

a_1, a_2 SPEC A $\quad :\gamma_4, \gamma_5, \tau_1, \tau_2, \delta_1, \delta_2$

$d_1(A) = r_1 \Longleftrightarrow d_2(A) = r_2 \quad :\gamma = f_\gamma(\gamma_i, \mu_i, \phi_i, \tau_i, \delta_i),$

$\alpha = f_\alpha(\gamma_i, \mu_i, \phi_i, \tau_i, \delta_i), \beta = f_\beta(\gamma_i, \mu_i, \phi_i, \tau_i, \delta_i)$

$d_1(A) \longleftrightarrow d_2(A) \quad :\gamma = f_\gamma(\gamma_i, \mu_i, \phi_i, \tau_i, \delta_i),$

$\alpha = f_\alpha(\gamma_i, \mu_i, \phi_i, \tau_i, \delta_i), \beta = f_\beta(\gamma_i, \mu_i, \phi_i, \tau_i, \delta_i)$

grain(Japan) = rice $\quad :\gamma$ = high, μ_r = high, μ_a = high, ϕ = high

continent(Japan) = Asia $\quad :\gamma$ = high, μ_r = low, μ_a = high, ϕ = moderate

grain(China) = rice $\quad :\gamma$ = high, μ_r = high, μ_a = high, ϕ = high

continent(China) = Asia $\quad :\gamma$ = high, μ_r = low, μ_a = high, ϕ = high

Japan, China SPEC place $\quad :\gamma_1, \gamma_2$ = high, τ_1, τ_2 = high, δ_1, δ_2 = low

---

grain(place) = rice $\Longleftrightarrow$ continent(place) = Asia

$:\gamma$ = low, α = low, β = moderate

grain(place) $\longleftrightarrow$ continent(place)

$:\gamma$ = moderate, α = low, β = moderate

(2) Generalization based on contrasting features

$d_1(a_1) = r_1$ $\quad :\gamma_1, \mu_{r1}, \mu_{a1}, \phi_1$

$d_2(a_1) = r_2$ $\quad :\gamma_2, \mu_{r2}, \mu_{a2}, \phi_2$

$d_1(a_2) = r_3$ $\quad :\gamma_3, \mu_{r3}, \mu_{a3}, \phi_3$

$d_2(a_2) = r_4$ $\quad :\gamma_4, \mu_{r4}, \mu_{a4}, \phi_4$

r_1 DIS r_3 $\quad :\gamma_5, \sigma_1$

r_2 DIS r_4 $\quad :\gamma_6, \sigma_2$

a_1, a_2 SPEC A $\quad :\gamma_7, \gamma_8, \tau_1, \tau_2, \delta_1, \delta_2$

$d_1(A) = r_1 \Longleftrightarrow d_2(A) = r_2 \quad :\gamma = f_\gamma(\gamma_i, \mu_i, \phi_i, \tau_i, \delta_i, \sigma_i),$

$\alpha = f_\alpha(\gamma_i, \mu_i, \phi_i, \tau_i, \delta_i, \sigma_i), \beta = f_\beta(\gamma_i, \mu_i, \phi_i, \tau_i, \delta_i, \sigma_i)$

$d_1(A) = r_3 \Longleftrightarrow d_2(A) = r_4 \quad :\gamma = f_\gamma(\gamma_i, \mu_i, \phi_i, \tau_i, \delta_i, \sigma_i),$

$\alpha = f_\alpha(\gamma_i, \mu_i, \phi_i, \tau_i, \delta_i, \sigma_i), \beta = f_\beta(\gamma_i, \mu_i, \phi_i, \tau_i, \delta_i, \sigma_i)$

$d_1(A) \longleftrightarrow d_2(A) \quad :\gamma = f_\gamma(\gamma_i, \mu_i, \phi_i, \tau_i, \delta_i, \sigma_i),$

$\alpha = f_\alpha(\gamma_i, \mu_i, \phi_i, \tau_i, \delta_i, \sigma_i), \beta = f_\beta(\gamma_i, \mu_i, \phi_i, \tau_i, \delta_i, \sigma_i)$

grain(South China) = rice $\quad :\gamma$ = high, μ_r = low, μ_a = high, ϕ = high

temperature (South China) = warm $\quad :\gamma$ = high, μ_r = low, μ_a = high, ϕ = moderate

grain(North China) = wheat $\quad :\gamma$ = high, μ_r = moderate, μ_a = high, ϕ = moderate

temperature(North China) = cool $\quad :\gamma$ = high, μ_r = moderate, μ_a = high,
ϕ = moderate

rice DIS wheat $\quad :\gamma$ = high, σ = low

warm DIS cool $\quad :\gamma$ = high, σ = low

(continued)

South China, North China SPEC place :γ_1, γ_2 = high, τ_1, τ_2 = high, δ_1, δ_2 = low

---

grain(place) = rice $\Longleftrightarrow$ temperature(place) = warm
:γ = low, α = low, β = moderate
grain(place) = wheat $\Longleftrightarrow$ temperature(place) = cool
:γ = low, α = low, β = moderate
grain(place) $\longleftrightarrow$ temperature(place)
:γ = moderate, α = low, β = moderate

(3) Refining a dependency to form an implication

$d_1(A) \longleftrightarrow d_2(A)$:γ_1, α, β
$d_1(a) = r_1$:$\gamma_2, \mu_{r1}, \mu_{a1}, \phi_1$
$d_2(a) = r_2$:$\gamma_3, \mu_{r4}, \mu_{a5}, \phi_6$
a SPEC A :γ_4, τ, δ

$d_1(A) = r_1 \Longleftrightarrow d_2(A) = r_2$
:$\gamma = f_\gamma(\gamma_i, \mu_i, \phi_i, \tau, \delta)$,
$\alpha = f_\alpha(\gamma_i, \mu_i, \phi_i, \tau, \delta)$, $\beta = f_\beta(\gamma_i, \mu_i, \phi_i, \tau, \delta)$

grain(place) $\longleftrightarrow$ temperature(place) :γ = moderate, α = low, β = moderate
grain(Saskatchewan) = wheat :γ = high, μ = high, ϕ = moderate
temperature(Saskatchewan) = cool :γ = high, μ_r = moderate, μ_a = high, ϕ = moderate
Saskatchewan SPEC place :γ = high, τ = high, δ = low

---

gain(place) = wheat $\Longleftrightarrow$ temperature(place) = cool
:γ = moderate, α = low, β = moderate

(4) Refining an implication from negative evidence

$d_1(A) = r_1 \Longleftrightarrow d_2(A) = r_2$:$\nu_1, \gamma_1, \alpha, \beta$
$d_1(a) = r_1$:$\gamma_2, \mu_{r1}, \mu_{a1}, \phi_1$
$d_2(a) = r_3$:$\nu_2, \gamma_3, \mu_{r2}, \mu_{a2}, \phi_2$
a SPEC A :γ_4, τ, δ
r_3 DIS r_2 :γ_5, σ

$d_1(A) = r_1 \Longleftrightarrow d_2(A) = \{r_2, r_3\}$:$\gamma = f(\gamma_i, \mu_i, \phi_i, \tau, \delta, \sigma), \alpha, \beta$

$$\phi_1' = \frac{\nu_1\phi_1}{\nu_1 + \nu_2}, \phi_2' = \frac{\nu_2}{\nu_1 + \nu_2}$$

grain(place) = wheat $\Longleftrightarrow$ temperature(place) = cool
:γ = moderate, α = low, β = moderate
grain(Italy) = wheat: γ = high, μ_r = high, μ_a = high, ϕ = moderate
temperature(Italy) = mild: γ = high, μ_r = moderate, μ_a = high, ϕ = moderate
mild DIS cool :γ = high, σ = moderate
Italy SPEC place :γ = high, τ = high, δ = low

---

grain(place) = wheat $\Longleftrightarrow$
temperature(place) = {cool, ϕ_1' = .5; mild, ϕ_2' = .5}
:γ = moderate, α = low, β = moderate

(continued)

TABLE 10 *(Continued)*

(5) Refining an implication from positive evidence

$d_1(A) = r_1 \Longleftrightarrow d_2(A) = \{r_2, r_3\}$ $\quad :\phi_1, \phi_2, \nu_1, \gamma_1, \alpha, \beta$
$d_1(a) = r_1$ $\quad :\gamma_2, \mu_{r1}, \mu_{a1}, \phi_1$
$d_2(a) = r_2$ $\quad :\nu_2, \gamma_3, \mu_{r2}, \mu_{a2}, \phi_2$
a SPEC A $\quad :\gamma_4, \tau, \delta$

$d_1(A) = r_1 \Longleftrightarrow d_2(A) = \{r_2, r_3\}$ $\quad :\gamma = f(\gamma_i, \mu_i, \phi_i, \tau, \delta), \alpha, \beta$

$$\phi_1' = \frac{\phi_1\nu_1 + \nu_2}{\nu_1 + \nu_2}, \phi_2' = \frac{\phi_2\nu_1}{\nu_1 + \nu_2}$$

grain(place) = wheat $\Longleftrightarrow$
temperature(place) = {cool, $\phi = .5$; mild, $\phi = .5$}
$:\gamma$ = moderate, α = low, β = moderate
grain(North China) = wheat
$:\gamma$ = high, μ_r = high, μ_a = high, ϕ = moderate
temperature(North China) = cool
$:\gamma$ = high, μ_r = moderate, μ_a = high, ϕ = moderate
North China SPEC place $:\gamma$ = high, τ = high, δ = low

---

grain(place) = wheat $\Longleftrightarrow$
temperature(place) = {cool, $\phi = .7$; mild, $\phi = .3$}
$:\gamma$ = moderate, α = low, β = moderate

(6) Generalizing an implication by referent combination

$d_1(a) = r_1 \Longleftrightarrow d_2(a) = \{r_2, r_3 \ldots\}$ $\quad :\phi_1, \phi_2, \nu_1, \gamma_1, \alpha, \beta$
r_2, r_3 SPEC R $\quad :\gamma_2, \gamma_3, \tau_1, \tau_2, \delta_1, \delta_2,$

$d_1(a)$ 1 $r_1 \Longleftrightarrow d_2(a) = R$ $\quad :\gamma = f_\gamma(\gamma_i, \mu_i, \phi_i, \tau_i, \delta_i)$

grain(place) = rice $\Longleftrightarrow$
climate(place) = {subtropical, $\phi = .5$, tropical, $\phi = .3$}
$:\gamma$ = moderate, α = moderate, β = moderate
tropical, subtropical SPEC hot
$:\gamma_1, \gamma_2$ = high, τ_1, τ_2 = high, δ_1, δ_2 = moderate

---

grain(place) = rice $\Longleftrightarrow$ climate(place) = hot
$:\gamma$ = moderate, α = moderate, β = moderate

temperature of the place. These are somewhat more certain generalizations than the common-feature generalizations but only marginally so. In general, people want to adduce more evidence for either so that some threshold of certainty can be achieved at which they are willing to consider seriously such hypotheses.

The third rule was identified from the protocols where subjects instantiated dependencies in terms of the referents they identified for particular cases. In this way, the subjects formed 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).

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 produces wheat and which has a 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.

Rule 6 generalizes an implication by generalizing over the set of referents covered in the consequent (or antecedent) of the implication. For example, if places that grow rice are either tropical or subtropical, then one might concluded that such places generally are hot.

We think these generalization rules incorporate all the ways that subjects formed new statements in the experiments. However, we formulated the rules to be as general as possible. People often make generalizations based on what appears to be insufficient evidence, but they are constantly refining their generalizations and often rejecting them as too uncertain to take seriously. So, the rule set we developed will surely produce generalizations no one will believe if applied willy nilly. People prevent themselves from making inappropriate generalizations by using other knowledge inferentially; that is, they restrict their generalizations to beliefs consistent with what they know in general (see Collins & Michalski, 1989).

Reasoning With Ordered Variables and Inequalities

In analyzing the protocols from the experiment described in the previous section, we extended 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 also may 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 they develop approximate ranges on those scales that they can refer to by 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 of 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.

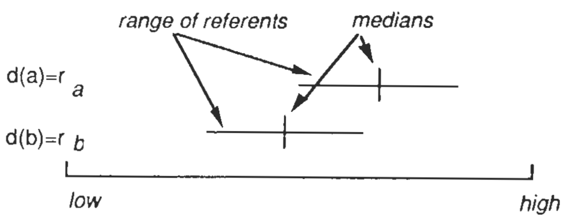
Ranges on ordered scales can be considered similar to the degree that they overlap. When one range overlaps the median or midpoint of another range and vice versa, then the two ranges can be considered highly similar. For ranges and values that are not highly similar, we introduce the inequality relations $<$, $>$, $\leq$, and $\geq$. Although we have not systematically studied how people interpret and compare these inexact ranges, for the sake of this article, we define these relations as follows: The statement $d(a) > r$ means that the referent of d(a) is from the midpoint of range r upward. The statement $d(a) \geq r$ means that the referent of d(a) is from the lower limit of range r, upward. Similarly, given $d(a) = r_a$ and $d(b) = r_b$, $d(a) > d(b)$ is equivalent to $r_a > r_b$, which means that r_a is dissimilar to r_b because the median value of r_a is greater than the upper limit value of r_b and the bottom value of r_a is greater than the median value of r_b (see Figure 1).

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 $<$, $>$, $\leq$, and $\geq$. For example, in a SIM-based argument transform, we rewrite the rule as:

$d(a_1) \sim r$
a_2 SIM a_1 in CX(A, D)
$D(A) \longleftrightarrow d(A)$
a_1, a_2 SPEC A

$d(a_2) \sim r$ where $\sim$ was one of $=$, $<$, $>$, $\leq$, or $\geq$.

In addition to this slight reformulation of the inference rules of the core theory, the introduction of inequalities yields some new generalization,



d(a) > d(b)

FIGURE 1 Inequality between continuous ranges.

transformation, and derivation inferences involving these orderings. All 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) 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 11 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 by 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 11), the two arguments a_1 and a_2 are related by an inequality instead of by similarity [as by $d_1(a_1) < d_1(a_2)$]. This makes sense because the notion of comparison is simply being extended to encompass comparison on an ordinal scale.

These rules extend to cases where several factors together determine some descriptor d_3 (by either an additive or conjunctive dependency). In this kind of generalization, it is assumed that other things are held constant. For example, with Table 11, 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 as a result of the reduced similarity of the two arguments. Rules 1 and 2 in Table 11 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 are ordered in the reverse direction for a negative dependency.

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 11 show these inference patterns. For example, Rule 3 can be used with the dependency between precipitation, rivers, and water supply discussed in the section A Sample Protocol: If places with light precipitation and a river have a moderate water supply (as described for Egypt and Italy), then one can conclude that a place like France with rivers and greater amounts of precipitation has a greater overall water supply because of the directed dependency bearing on water supply. 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 11

Inequality-Generating Inferences

Inequality transforms with directed dependencies

$d_1(A) \overset{+}{\longleftrightarrow} d_3(A)$ $:\gamma_1, \alpha_1$

$d_1(A)$ & $d_2(A)^* \longleftrightarrow d_3(A)$

$:\gamma_2, \gamma_3, \alpha_2, \alpha_3$

a_1, a_2 SPEC A: $\gamma_4, \gamma_5, \tau_1, \tau_2$

a_2 SIM a_1 in CX(A, $d_2(A)^*$): γ_6, σ_1

$d_1(a_1) = r_1 :\gamma_7, \mu_{r1}, \mu_{a1}$

$d_1(a_2) \sim r_1 :\gamma_8, \mu_{r2}, \mu_{a2}$

$d_3(a_1) = r_3 :\gamma_9, \mu_{r3}, \mu_{a3}$

$d_3(a_2) \sim r_3 :\gamma = f_\gamma(\gamma_i, \mu_i, \alpha_i, \tau_i, \sigma_i)$

$d_1(A) \overset{-}{\longleftrightarrow} d_3(A)$ $:\gamma_1, \alpha_1$

$d_1(A)$ & $d_2(A)^* \longleftrightarrow d_3(A)$

$:\gamma_2, \gamma_3, \alpha_2, \alpha_3$

a_1, a_2 SPEC A: $\gamma_4, \gamma_5, \tau_1, \tau_2$

a_2 SIM a_1 in CX(A, $d_2(A)^*$): γ_6, σ_1

$d_1(a_1) = r_1 :\gamma_7, \mu_{r1}, \mu_{a1}$

$d_1(a_2) \sim r_1 :\gamma_8, \mu_{r2}, \mu_{a2}$

$d_3(a_1) = r_3 :\gamma_9, \mu_{r3}, \mu_{a3}$

$d_3(a_2) \sim^- r_3 :\gamma = f_\gamma(\gamma_i, \mu_i, \alpha_i, \tau_i, \sigma_i)$

Notes: $d_2(A)^*$ stands for all other terms that $d_3(A)$ depends on.

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

altitude(place) $\overset{-}{\longleftrightarrow}$ temperature(place) in CX(places, latitude)

$:\gamma$ = high, α = moderate

altitude(place) & latitude(place) $\longleftrightarrow$ temperature(place)

$:\gamma$ = high, α = high

Miami, Mexico City SPEC place: γ = high, τ = high, δ = low

Miami SIM Mexico City in CX(places, latitude): γ = high, σ = moderate

altitude(Mexico City) = high: γ = high, μ_r = low, μ_a = high

altitude(Miami) < high: γ = high, μ_r = low, μ_a = high

temperature(Mexico City) = moderate: γ = high, μ_r = low, μ_a = high

---

temperature(Miami) > moderate: γ = high

Derivations from implications with directed dependencies

$d_1(A) \overset{+}{\longleftrightarrow} d_3(A)$ $:\gamma_1, \alpha_1$

$d_1(A) = r_1$ & $d_2(A) = r_2^*$ $\Longleftrightarrow d_3(A) = r_3$ $:\gamma_2, \gamma_3, \alpha_2, \alpha_3$

a SPEC A $:\gamma_4, \tau_1, \delta$

$d_2(a) = r_2$ $:\gamma_5, \mu_{r1}, \mu_{a1}$

$d_1(a) \sim r_1$ $:\gamma_6, \mu_{r2}, \mu_{a2}$

$d_3(a) \sim r_3$ $:\gamma = f_\gamma(\gamma_i, \mu_i, \alpha_i, \tau_i, \sigma_i)$

$d_1(A) \overset{-}{\longleftrightarrow} d_3(A)$ $:\gamma_1, \alpha_1$

$d_1(A) = r_1$ & $d_2(A) = r_2^*$ $\Longleftrightarrow d_3(A) = r_3$ $:\gamma_2, \gamma_3, \alpha_2, \alpha_3$

a SPEC A $:\gamma_4, \tau_1, \delta$

$d_2(a) = r_2$ $:\gamma_5, \mu_{r1}, \mu_{a1}$

$d_1(a) \sim r_1$ $:\gamma_6, \mu_{r2}, \mu_{a2}$

$d_3(a) \sim^- r_3$ $:\gamma = f_\gamma(\gamma_i, \mu_i, \alpha_i, \tau_i, \sigma_i)$

Notes: $d_2(A) = r_2^*$ stands for all other terms in the implication.

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

precipitation(place) & has-river(place) $\overset{+}{\longleftrightarrow}$ water-supply(place)

$:\gamma_1, \gamma_2$ = high, α_1, α_2 = high

precipitation(place) = very light & has-river(place) = yes

$\Longleftrightarrow$ water-supply(place) = moderate: γ_1, γ_2 = high, α_1, α_2 = high

France SPEC place: γ = high, τ = high, δ = low

has-river(France) = yes: γ = high, μ_r = low, μ_a = high

precipitation(France) > very light: γ = high, μ_r = low, μ_a = high

---

water-supply(France) > moderate: γ = high

As previously mentioned, both of these inference forms have the requirement that other relevant "contextual factors" are held constant. In the dependency-based transform rules (Table 11, 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 $a_3(A)$ [represented by $d_2(A)^* d_3(A)$]. 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 [& $d_2(A) = r_{2*} \ldots$] 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.

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 9, Rule 2). This is shown in Table 12 for the simple case of comparing attributes of two exemplars. As an example of this type of generalization (Table 12, 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. Similar rules can be used to refine an unspecified dependency, essentially using the dependency to select the attributes that need to be compared. For instance, the generalization from Alaska and Equador would be made more certain if it were known beforehand that a dependency existed between latitude and temperature but that the form of the relationship were unknown until the exact data were considered.

Another way to derive a directed dependency is by using two implications that address the same pair of descriptors. These generalization rules are similar to 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 12. As in the previous example, the referents of the corresponding terms in the implications are placed in correspondence and their values compared. Because the direction of shift from a tropical to a polar place is opposite on the two descriptors, a negative dependency is formed.

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 have yet to be resolved, some of which are touched on in the next section. For example, there is a trade-off between the precision and certainty of a referent value that pervades this kind of reasoning. We do not yet

TABLE 12
Generalizing Based on Inequalities

Generalizing to specified dependencies

$d_1(a_1) = r_1: \gamma_1, \mu_{r1}, \mu_{a1}$	$d_1(a_1) = r_1: \gamma_1, \mu_{r1}, \mu_{a1}$
$d_2(a_1) = r_2: \gamma_2, \mu_{r2}, \mu_{a2}$	$d_2(a_1) = r_2: \gamma_2, \mu_{r2}, \mu_{a2}$
$d_1(a_2) = r_3: \gamma_3, \mu_{r3}, \mu_{a3}$	$d_1(a_2) = r_3: \gamma_3, \mu_{r3}, \mu_{a3}$
$d_2(a_2) = r_4: \gamma_4, \mu_{r4}, \mu_{a4}$	$d_2(a_2) = r_4: \gamma_4, \mu_{r4}, \mu_{a4}$
a_1, a_2 SPEC A: $\gamma_5, \gamma_6, \tau_1, \tau_2, \delta_1, \delta_2$	a_1, a_2 SPEC A: $\gamma_5, \gamma_6, \tau_1, \tau_2, \delta_1, \delta_2$
$r_1 \sim r_3: \gamma_7$	$r_1 \sim r_3: \gamma_7$
$r_2 \sim r_4: \gamma_8$	$r_2 \sim^{-} r_4: \gamma_8$
$d_1(A) \overset{+}{\longleftrightarrow} d_2(A)$	$d_1(A) \overset{-}{\longleftrightarrow} d_2(A)$

$:\gamma = f_\gamma(\gamma_i, \mu_i, \tau_i, \delta_i),\ \alpha = f_\alpha(\gamma_i, \mu_i, \tau_i, \delta_i),\ \beta = f_\beta(\gamma_i, \mu_i, \tau_i, \delta_i)$

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

latitude(Alaska) = high $:\gamma$ = high, μ_r = low, μ_a = high
temperature-range(Alaska) = cold: γ = high, μ_r low, μ_a = high
latitude(Equador) = low $:\gamma$ = high, μ_r low, μ_a = high
temperature-range(Equador) = hot: γ = high, μ_r low, μ_a = high
Equador, Alaska SPEC place $:\gamma_1, \gamma_2$ = high, τ_1, τ_2 = low, δ_1, δ_2 = low
hot > cold $:\gamma$ = high
low < high $:\gamma$ = high

---

latitude(place) $\overset{-}{\longleftrightarrow}$ temperature-range(place)
$:\gamma$ = low, α = moderate, β = low

Generalizing implications to directed dependencies

$d_1(A) \longleftrightarrow d_2(A): \gamma_1, \alpha_1, \beta_1$	$d_1(A) \longleftrightarrow d_2(A): \gamma_1, \alpha_1, \beta_1$
$d_1(a_1) = r_1 \Longleftrightarrow d_2(a_1) = r_2$ $:\gamma_2, \alpha_2, \beta_2$	$d_1(a_1) = r_1 \Longleftrightarrow d_2(a_1) = r_2$ $:\gamma_2, \alpha_2, \beta_2$
$d_1(a_2) = r_3 \Longleftrightarrow d_2(a_2) = r_4$ $:\gamma_3, \alpha_3, \beta_3$	$d_1(a_2) = r_3 \Longleftrightarrow d_2(a_2) = r_4$ $:\gamma_3, \alpha_3, \beta_3$
a_1, a_2 SPEC A: $\gamma_4, \gamma_5, \tau_1, \tau_2, \delta_1, \delta_2$	a_1, a_2 SPEC A: $\gamma_4, \gamma_5, \tau_1, \tau_2, \delta_1, \delta_2$
$r_1 \sim r_3: \gamma_6$	$r_1 \sim r_3: \gamma_6$
$r_2 \sim r_4: \gamma_7$	$r_2 \sim^{-} r_4: \gamma_7$
$d_1(A) \overset{+}{\longleftrightarrow} d_2(A)$	$d_1(A) \overset{-}{\longleftrightarrow} d_2(A)$

$:\gamma = f_\gamma(\gamma_i, \mu_i, \tau_i, \delta_i),\ \alpha = f_\alpha(\gamma_i, \mu_i, \tau_i, \delta_i),\ \beta = f_\beta(\gamma_i, \mu_i, \tau_i, \delta_i)$

latitude(place) $\longleftrightarrow$ temp-range(place) $:\gamma$ = low, α = moderate, β = low
latitude(tropical-place) = low $\Longleftrightarrow$ temp-range(tropical-place) = hot
$:\gamma$ = low, α = high, β = high
latitude(polar-region) = high $\Longleftrightarrow$ temp-range(polar-region) = cold
$:\gamma$ = low, α = high, β = low
tropical-place, polar-region SPEC place
$:\gamma_1, \gamma_2$ = high, τ_1, τ_2 = low, δ_1, δ_2 = low
hot > cold $:\gamma$ = high
low < high $:\gamma$ = high

---

latitude(place) $\overset{-}{\longleftrightarrow}$ temperature-range(place)
$:\gamma$ = low, α = high, β = moderated

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.

CONCLUSION

This article has described some revisions to the core theory of human plausible inference developed by Collins and Michalski (1989). As in that earlier article, our work is aimed at formalizing the plausible inferences observed in people's answers to questions for which they do not have ready answers. Both the theory described in this article and our discussion of some as yet unresolved issues were motivated in large part by a protocol experiment that was designed to bring forth more clearly people's uses of multiple sources of evidence in forming plausible conclusions. We present examples of how this experimental evidence supported the introduction of a mechanism for reasoning about inequalities and a reanalysis of backward reasoning in the presence of multiple dependencies.

Another important result of the experiment was a clarification of the need for an integrated theory of generalization to accompany the core plausible reasoning theory. This was the subject of the second half of the article, which included a number of patterns of plausible generalization and generalization refinement, including rules for reducing the certainty of generalizations in the presence of contradictory evidence.

In the revised theory, we addressed only those issues raised by the experiment that we could find solutions for. There are a number of other issues in the experiment and earlier protocols that we have not addressed. We think they are amenable to the kind of analysis we have been using, but the solutions were not readily apparent or we lacked the time to pursue them. We enumerate them so that the reader can see what 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 backward or forward over dependencies and implications where a number of variables (e.g., precipitation and rivers) affected a particular variable (e.g., water supply). Subjects treated the variables as if they were ORed together, 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 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 trade-off 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 in. 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 trade-offs 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 temperature of the Amazon jungle averages 85 °F, or 1 mi in altitude affects temperature as much as 800 mi in latitude). There needs to be a smooth method of incorporating 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 article, we did specify how the numeracy parameter ν can logically be combined to derive frequency ϕ. It should be possible to develop a normative theory that combines all the parameters specified in the theory; however, 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 versus parades versus 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, certain internal organs are found in a wider range of animals than are horns or colors. Therefore, a person is more likely to infer that an animal has a gizzard because a similar animal has one than to infer that an animal has a horn because a similar animal has one. We have not incorporated 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. The data from the experiment suggested that each piece of information that is found redirects the search process in memory. Therefore, we think 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 are not yet ready to tackle.

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

ACKNOWLEDGMENTS

Research for this article 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.

REFERENCES

Baker, M., Burstein, M. H., & Collins, A. (1987). Implementing a model of human plausible reasoning. In J. McDermott (Ed.), *Proceedings of the Tenth International Joint Conference of Artificial Intelligence, 1* (pp. 185-188). Los Altos, CA: Kaufmann.

Burstein, M. H., & Collins, A. (1988). Modeling a theory of human plausible reasoning. In T. O'Shea & V. Sgurev (Eds.), *Artificial intelligence 3: Methodology, systems, applications* (pp. 21-28). Amsterdam: North Holland.

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

Cohen, P. R. (1987). The control of reasoning under uncertainty: A discussion of some programs. *The knowledge engineering review* (Vol. 2, No. 1). Cambridge University Press.

Cohen, P. R., & Loiselle, C. L. (1988). Beyond ISA: Structures for plausible inference in semantic networks. In *Proceedings of the Seventh National Conference on Artificial Intelligence.* St. Paul, MN: Kaufmann.

Collins, A. (1978). Fragments of a theory of human plausible reasoning. In D. Waltz (Ed.), *Theoretical issues in natural language processing 2* (pp. 194-201). Urbana, IL: University of Illinois Press.

Collins, A., & Stevens, A. L. (1982). Goals and strategies of inquiry teachers. In R. Glaser (Ed.), *Advances in instructional psychology II.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

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

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

DeJong, G., & Mooney, R. (1986). Explanation-based learning: An alternative view. *Machine Learning, 1,* 145–176.

Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Ann. Math Statistics, 38,* 325–339.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1987). *Induction.* Cambridge, MA: MIT Press.

Huhns, M. N., & Stephens, L. M. (1989). Plausible inferencing using extended composition. In N. S. Sridharan (Ed.), *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence* (pp. 1420–1425). Los Altos, CA: 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 P. Langley (Ed.), *Proceedings of the Fourth International Workshop on Machine Learning* (pp. 50–58). Los Altos, CA: 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 A. Bundy (Ed.), *Proceedings of the 8th International Joint Conference on Artificial Intelligence* (pp. 1139–1151). Los Altos, CA: Kaufmann.

Mitchell, T. M., Keller, R. M., and Kedar-Cabelli, S. T. (1986). Explanation-based generalization: A unifying view. *Machine Learning, 1*(1), 47–80.

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

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, CA: Kaufmann.

Pearl, J. (1990). Bayesian and belief-functions formalisms for evidential reasoning: A conceptual analysis. In G. Shafer & J. Pearl (Eds.), *Reading on uncertain reasoning* (pp. 540–574). San Mateo, CA: Kaufmann.

Russell, S. J. (1989). *The use of knowledge in analogy and induction.* London: Pitman.

Shafer, G. (1976). *A mathematical theory of evidence.* Princeton, NJ: Princeton University Press.

Smith, E. E., & Osherson, D. N. (1984). Conceptual combination with prototype concepts. *Cognitive Science, 8,* 337–361.

Smith, E. E., & Osherson, D. N. (1989). Similarity and decision making. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 60–75). Cambridge, England: Cambridge University Press.

Vosniadou, S., & Ortony, A. (Eds.). (1989). *Similarity and analogical reasoning.* Cambridge, England: Cambridge University Press.

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