Extending a Model of Human Plausible Reasoning

Mark H. Burstein
Allan M. Collins*
Bolt Beranek and Newman Inc.
10 Moulton St.
Cambridge, MA 02138
EMAIL: BURSTEIN@BBN.COM
* also the Institute for the Learning Sciences

Abstract

When one looks at transcripts of people answering questions or carrying on dialogues about everyday matters, their comments are filled with plausible inferences -- inferences that are not certain, but that make sense. Often, in forming these inferences, generalizations are made that are equally uncertain, but are nevertheless useful as a guide to their reasoning. This paper describes some extensions to our earlier description of a core theory of plausible reasoning (Collins and Michalski, 1989), based in large part on a recent protocol study. The primary focus is on the inductive inference patterns people use to form *plausible generalizations*, weakly held beliefs based on few examples. We also show how the model was extended to deal with plausible inferences involving continuous quantities and inequalities.

Introduction

In a previous paper (Collins & Michalski, 1989) we presented a core theory of human plausible reasoning based on analysis of a large set of people's answers to everyday questions about the world (Collins 1978). The theory was characterized in terms of Michalski's (1987) variable-valued logic notation. This paper briefly describes several kinds of revisions to the formalism developed in that earlier paper, based on the results of a more recent protocol study.

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

The other two types of expressions in Table 1 are called mutual implications and mutual dependencies.A mutual implication specifies how one statement is related to another statement. The example states that warm temperature and heavy rainfall imply rice growing, and vice versa. A mutual dependency relates two *terms* (e.g. latitude(place) and temperature(place)). Our model of dependencies and their use in plausible inferences is much like Russell's(1986) notion of determinations. The example represents the belief that the average temperature of a place is inversely related to its latitude.

Each type of statement carries with it various kinds of certainty information that affects inference conclusion certainty :

- **Certainty** (γ) annotates all state- ments. It indicates overall level of belief in a statement.
- ***Typicality*** (τ) annotates GEN and SPEC statements. The more typical England is of Europe, or Surrey is of England, with respect to climate (or any variable that affects flower growing), the more certain the inference.
- ***Similarity*** (σ) annotates the SIM and DIS statements. Hence, the more similar Holland is to England, and the less similar Java is to England, with respect to climate, the more certain the inference.
- ***Conditional Likelihood*** (α,β) modifies dependencies and implications. For example, α indicates how much latitude depends on temperature, and β how much temperature depends on latitude.

  Frequency (ϕ) reflects the all/some distinction in logic, but as a continuous variable. I.e., 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.

Table 1: Different Types of Expressions in the Core Theory

	Form	Examples
Statements (S):	$d(a) = r$	means-of-locomotion(birds)={flying...}
Relational Statements (R)	a_1 REL a_2 in CX (A,d)	
	a_1 GEN a_2 in CX (A,d)	bird GEN robin in CX (birds, all characteristics)
	a_1 SPEC a_2 in CX (A,d)	chicken SPEC fowl in CX (birds, biological chars.)
	a_1 SIM a_2 in CX (A,d)	duck SIM goose in CX (birds, habitat)
	a_1 DIS a_2 in CX (A,d)	duck DIS goose in CX (birds, neck length)
Mutual Implications (I)	$d_1(a) = r_1$ <==> $d_2(a) = r_2$	temperature(place)=warm & rainfall(place)=heavy <==> grain(place)=rice
Mutual Dependencies (D)	$d_1(a)$ <--±--> $d_2(a)$	average temperature (place) <--⁻--> latitude (place)

- ***Dominance*** (∂) applies to GEN and SPEC inferences and reflects the degree the subset comprises a large part of the set. For example, since Surrey in only a small part of England, growing daffodils and roses is less certain than for Southern England as a whole.
- ***Multiplicity of the argument*** (μ_a) reflects the degree to which more than one country (the superordinate of the argument) has daffodils and roses. Since many countries presumably have daffodils and roses, μ_a is high and the argument-based inferences are more certain (except for the DIS inference).
- ***Multiplicity of the referent*** (μ_r) reflects the degree to which England has more types of flowers (the superordinate of the referent) than daffodils and roses. Since countries usually have many different types of flowers, μ_r is high and the referent-based inferences more certain (except for the DIS inference).

A *similarity transform* (a basic inference type) would make use of these as follows:

flower-type(England) = roses
  : γ = high, μ_a = high, μ_b = high, ϕ = high
England SIM Holland in CX (place, climate)
  : γ = high, σ = high
climate(place) <---> flower-type(place)
  : γ = high, α = high, β = moderate
England, Holland SPEC place
  : γ = high, τ = high, δ = low
___
flower-type(Holland) = roses: : γ = high

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*); 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*). In general, when an inference involves a dependancy, relationships between terms (GEN, SPEC, SIM, DIF) these relationships are specified within a context, CX, that relates to the dependancy governing the inference. The overall certainty of the conclusion will vary depending on the similarity/difference between the related terms *in that context*, not simply with respect to overall similarity (which is written as "all characteristics").

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

On the right side of the table, is a matrix showing how and in which direction each certainty parameter affects each of these transforms. Thus, typicality (τ), which relates terms to their specializations, affects GEN and SPEC transforms, while similarity (σ) affects SIM and DIS transforms. 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 parameters are described in more detail in the Collins & Michalski (1989).

There are a number of other inference types over these forms of statements that are described in detail in Collins and Michalski (1989), and a computer implementation of the core theory was described in Baker et al (1987).

Table 2: Eight Transforms on the Statement "flower-type(England)={daffodils, roses...}"

	Parameter:	τ	σ	α	β	δ	μ_a	μ_r
Argument-based Transforms								
(1) **GEN**	flower-type(Europe)={daffodils, roses...}	+	0	+	+	+	+	0
(2) **SPEC**	flower-type(Surrey)={daffodils, roses...}	+	0	+	+	+	0	0
(3) **SIM**	flower-type(Holland)={daffodils, roses...}	0	+	+	+	0	+	0
(4) **DIS**	flower-type(Java)={daffodils, roses...}	0	-	+	-	0	-	0
Referent-based Transforms								
(5) **GEN**	flower-type(England)={temperate flowers...}	+	0	+	+	+	0	+
(6) **SPEC**	flower-type(England)={yellow roses...}	+	0	+	+	+	0	0
(7) **SIM**	flower-type(England)={peonies...}	0	+	+	+	0	0	+
(8) **DIS**	flower-type(England)={bougainvillea...}	0	-	+	-	0	0	-

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

A Protocol Experiment

The extensions to the core theory of Collins and Michalski (1989) described in this paper 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 to unknown information. In this experiment, designed and conducted by Michell Baker, we gave human subjects a partially-specified matrix of geographical variables crossed by countries, shown in Table 3, that we had developed for the computer simulation. We then interviewed five different scientists (who were not geographers) as they attempted to fill in the missing cells in the matrix. 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.

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 one half hour to one and one half hours for different subjects. Each session was recorded on audio tape and transcribed.

While the tasks 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 specified in earlier papers (Collins 1978; Collins and 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 made it different from earlier protocols we had collected. Because the subjects had the matrix in front of them, they could consider more cases and variables than the one or two they usually do. Second, because 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. So these protocols had more 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.

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

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; Mediteraneans 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.

Table 3: Experimental Matrix of Geographic Data

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

In the first part of this response, the subject focused on the evidence that the Italian climate was Mediterranean, and the fact that there were mountains. The Mediterranean climate led the subject to infer that Italy had limited precipitation, while the presence of mountains indicated that there would be more rain than other, similar, lowland areas with the same general climate. In the second half of the response, the subject based his inferences on the evidence that the fresh water supply given for Italy was moderate, and the fact that there were rivers. There are two kinds of uncertain information here. One is the question of how much of a contribution each of those factors can make to overall water supply, a question for which the subject presumably had little direct knowledge. The other problem is the lack of information of even a qualitative kind as to the amount of water available from rivers. All the matrix supplied was the fact that there were at least *some* rivers.

It appears from the pattern of this subject's protocols, and from subsequent questioning of the subject, that he normally treated water supply as if it was directly dependent on precipitation, independently of the presence of rivers. Other subjects treated the two as more "additive". In general, either precipitation or rivers could account for the water supply of a place, and this subject generally assumed water supply was directly correlated with precipitation, unless there was evidence to the contrary. In making a comparison to Egypt, the subject reasoned 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**.

Refining the Core Theory

Several issues were raised in this protocol that were not specifically covered by the core theory. The first was the reasoning about inequalities on ordinal referents. Another, related issue was the use of reasoning about deviation from a default value.The logic of this inference was: Whatever is determined to be the normal value of rainfall in a place based on other variables, mountains tend to make the rainfall higher. So if Mediterranean climates have light rainfall, the mountains would make the rainfall greater than light. A third problem for the core theory occurred when the "counter evidence" to this was considered. When the subject saw that Italy had rivers, and that accounted for Italy's moderate water supply (by analogy to Egypt), then that decreased the certainty of the inference that Italy's moderate water supply implied moderate precipitation. 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 can influence a dependent variable. The conclusion is not that the original inference was wrong, but that the evidence used to make the inference could be accounted for by other means.

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

> **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.

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
Water-supply(Italy) = moderate
____
Has-river(place) = yes & Precipitation(place) = light
    <==> Water-supply(place) = moderate

This generalization is one of the new rules added to the theory, described in detail in (Burstein, Collins and Baker, 1991). It is called **Refining a dependency to form an implication** (Table 10, Rule 3). Generalizations like above inference, 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 *explanation-based generalizations* described by DeJong (1981) and Mitchell (1983), except that the initial knowledge is in the form of dependencies, not full causal or inferential rules. All of these generalization mechanisms hinge on the combination of general causal or explanatory background knowledge with a new specific case or cases to form a new, potentially more useful general rule. We call the class of such generalizations *refinements.*

In this second protocol, the subject reasons that places with rivers and a little rainfall have moderate water supply, so places with abundant water supply must get more rain. Louisiana's abundant water supply, being greater than both Egypt and Italy's, means that it should have greater precipitation as well. The conclusion was that if Louisiana had an abundant water supply, then it must have "at least moderate" precipitation. In the formalization of this inference, we have described the pattern as predicting simply that it should be "greater than light", light being the corresponding value in the implication. We take these as equivalent with respect to a {low, medium, high} scale. (For precipitation, the term **light** corresponds to the more neutral referent **low**, and **abundant** is the same as **high**.)

Has-river(place) & Precipitation(place) <--+-->
    Water-supply(place)
Has-river(place) = yes & Precipitation(place) = light
    <==> Water-supply(place) = moderate
Has-river(Louisiana) = yes
Water-supply(Louisiana) = abundant (or > moderate)
____
Precipitation(Louisiana) > light (or $\geq$ moderate)

This is an example of a class of inferences that was not explicitly dealt with previously. The issue is reasoning with inequalities on continuous or ordered variables, in conjunction with dependencies between those types of variables. These inferences all depend on the presence of a *specified* dependency: specified dependencies are those labeled with a + or - to indicate that an increase/decrease in the values for a term on one side has a corresponding positive or negative effect on the other. However, by extending the formalism to include explicit orderings on referents and allowing all inferences that contain statements of the form d(a) = r to also allow the = to be replaced by any of <, >, $\leq$, and $\geq$, we were able to rewrite many of the original rules to cover this kind of inference (Burstein, Collins and Baker, 1991). For example, in a SIM-based argument transform, we would rewrite the rule as:

$d(a_1) \sim r$ where $\sim$ is one of =, < , >, $\leq$, or $\geq$.
a_2 SIM a_1 in CX (A, D)
D(A) <-----> d(A)
a_1,a_2 SPEC A
____
$d(a_2) \sim r$

This reformulation to introduce inequalities also yields some new derivation, transformation and generalization rules. All of the new inferences involve specified dependencies (dependencies of the form $d_1(A)$<--+-->$d_2(A)$ or $d_1(A)$<-- - -->$d_2(A)$), like the inference about precipitation above. The new rules for *creating* inequalities with SIM-based transforms on dependencies and implications are much like the SIM-based referent transform rules described in Collins and Michalski (1989). In the new rules for generating inequalities from directed dependencies, the two arguments a_1 and a_2 are related by an inequality instead of by similarity (as by $d_1(a_1) < d_1(a_2)$). For example, the relationship between latitude and temperature range might be inferred as follows:

latitude(Alaska) = high : γ = hi, μ_r = lo, μ_a = hi
temp-range(Alaska) = cold: γ = hi, μ_r = lo, μ_a = hi
latitude(Equador) = low: γ = hi, μ_r = lo, μ_a = hi
temp-range(Equador) = hot: γ = hi, μ_r = lo, μ_a = hi
Equador, Alaska SPEC place: γ_i = hi, τ_i = lo, δ_i = lo
hot > cold: γ = high
low < high: γ = high

latitude(place) <--˜--> temp-range(place)
: γ = low, α = moderate, β = low

Open Issues

The revisions to the core theory of human plausible inference developed by Collins and Michalski (1989), as described here and in (Burstein, Collins and Baker, 1991), were part of our continuing attempt to formalize the plausible inferences observed in people's answers to questions for which they do not have ready answers. The recent work was 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 have shown how this experimental evidence supported the introduction of a mechanism for reasoning about inequalities, and clarified the need for an integrated theory of generalization to accompany the core plausible reasoning theory.

There are a number of other issues apparent to us in the experiment and earlier protocols that we have not yet addressed. We think they are amenable to the kind of analysis we have been using, but the solutions were not as apparent. We list some of them here:

- Reasoning with complex dependencies and implications. Subjects frequently reasoned backwards over dependencies where a number of variables (e.g. precipitation and rivers) affected a particular variable (e.g. water supply). Sometimes these were treated as disjunctive variables, sometimes as conjunctive (neither alone was predictive), and sometimes as if their effects were additive. This may require several alternative mechanisms be introduced into the model, based on the form of the causal knowledge used.

- The tradeoff between range and certainty. Subjects appear to trade off certainty about a belief against the range of the referent. For example, one might be very certain that the average rainfall in Louisiana is "at least moderate," somewhat less certain that it is "heavy," still less certain that it is between 40 and 60 inches a year.

- Merging of qualitative and quantitative reasoning. Sometimes subjects bring in various quantitative relationships to guide their qualitative reasoning (e.g. the Amazon jungle averages 85° or 1 mile in altitude affects temperature as much as 800 miles in latitude).

- The extent parameter. Collins (1978) identified a parameter he called "extent" which was particularly prevalent in temporal and spatial inferences. It is necessary because people have a notion of how far rainstorms vs. parades vs. continents extend in space, and how long they extend in time. This notion is central to people's reasoning about space and time, and features of categories.

References

Baker, M., Burstein, M. H., and Collins, A. 1987. Implementing a model of human plausible reasoning. In *Proceedings of the Tenth International Joint Conference of Artificial Intelligence, 1*, 185-188. Morgan Kaufmann.

Burstein, M. H., Collins, A. and Baker, M., 1991. Plausible Generalization: Extending a Model of Human Plausible Reasoning. *Journal of the Learning Sciences.* (Forthcoming.)

Collins, A. 1978. Fragments of a theory of human plausible reasoning. In D. Waltz ed., *Theoretical Issues in Natural Language Processing II* 194-201. Urbana, IL: University of Illinois.

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

DeJong, G. and Mooney, R. 1986. Explanation-Based Learning: An Alternative View. *Machine Learning* 1:145-176.

Mitchell, T. M., Keller, R. M. and Kedar-Cabelli, S. T. 1986. Explanation-Based Generalization: A Unifying View. *Machine Learning* 1:1.

Pearl, J. 1987. Embracing causality in formal reasoning. In *Proceedings of the Sixth National Conference on Artificial Intelligence* 369-373. Morgan Kaufmann.

Russell, S. J. 1986. *Analogical and Inductive Inference.* Ph.D. thesis, Dept. of Computer Science, Stanford University.

Acknowledgements

We would like to thank Michelle Baker for her contributions to this work and her comments on an earlier draft of this paper.

This research was supported by the Army Research Institute under Contract No. MDA 903-85-C-0411.