

# Categorization vs. logic in category conjunction: towards a model of overextension

Fintan J. Costello (Fintan.Costello@ucd.ie)

Department of Computer Science, University College Dublin,  
Belfield, Dublin 4, Ireland.  
fintan.Costello@ucd.ie

## Abstract

Traditional approaches to semantics give a logical, set-theoretic account of category conjunction, where an item is a member of a conjunction  $A \& B$  only if it is a member of both single categories  $A$  and  $B$ . However, people do not always follow this logical approach when classifying items in conjunctions. For example, Hampton (1988) found that people typically classify blackboards as non-members of the single category “furniture”, but as members of the conjunction “school furniture”. This paper describes a computational model intended to explain this “overextension” of conjunctive categories, and an experiment testing this model using conjunctions of controlled, laboratory-learned categories.

## Introduction

The ability to form new complex categories by the conjunction of existing categories is an important part of classification and reasoning. Any worthwhile AI model of categorization should be able to reason about conjunctive categories in a way that mirrors human reasoning about such conjunctions. Traditional approaches to semantics give a logical, set-theoretic, account of category conjunction, in which an item will be a member of a conjunction  $A \& B$  only if it is a member of the single category  $A$  and a member of the single category  $B$ . However, a number of studies (Chater, Lyon, & Myers, 1990; Hampton, 1988; Storms, De Boeck, Van Mechelen, & Ruts, 1998) show that people do not follow this simple logical approach when classifying items in certain natural-language conjunctions. Instead, for some conjunctions a reliable pattern of “overextension” occurs. In those conjunctions people will classify some items as members of the conjunction despite having classified those items as non-members of one or other constituent category of the conjunction (the conjunction is “overextended” beyond the membership of its constituent categories). For example, Hampton (1988) found that most participants in his experiments classified “blackboard” as a member of the conjunction “school furniture”, but judged “blackboard” a non-member of the single category “furniture”.

Explaining overextension is difficult for standard AI models of classification, which typically assume that an item’s degree of membership in a conjunction is obtained by taking the minimum (or sometimes the product) of its membership scores in the two constituent categories of that conjunction. On the face of it, overextension seems to be an example of people reasoning illogically about category membership and conjunction. However, in this paper I argue that overextension is a logical consequence of the

need to use both positive and negative evidence in judging membership in category conjunctions. I outline a computational model of conjunction that uses positive and negative evidence to produce overextension for some specific items, and describe an experiment testing the model by comparing its predictions to people’s classification of items in conjunctions of artificial, laboratory-learned categories.

## The Diagnostic Evidence Model

The diagnostic evidence model is an extension of a model originally developed to explain how people interpret novel noun-noun phrases (Costello & Keane, 2000). The model aims to explain classification in both single categories (see Costello, 2000) and category conjunctions (Costello, 2001). The model assumes that people represent categories by storing sets of category members in memory. From these sets, diagnostic attributes for categories are computed: these attributes serve to identify new category members. An item’s classification in a single category or in a conjunction is a function of the diagnosticity of its attributes for that category or for the constituent categories of that conjunction. An item has a high classification score in a category if it has diagnostic attributes of that category. An item has a high score in a conjunction if it has some attributes diagnostic for one constituent of the conjunction, and others diagnostic for the other.

## Attribute Diagnosticity

Diagnostic attributes are attributes which occur frequently in members of a category, but rarely in that category’s contrast set (the set of non-members of that category). These attributes serve to identify members of a category: a new item having an attribute which is diagnostic for a category is likely to be a member of that category. Equation 1 defines the diagnosticity of an attribute  $x$  for a category  $C$ . Let  $K$  be  $C$ ’s contrast set. Let  $j_x$  be 1 if an item  $j$  has attribute  $x$ , and 0 otherwise.  $D(x|C|K)$ , the diagnosticity of  $x$  for  $C$  relative to  $K$ , is equal to the number of members in  $C$  which have  $x$ , divided by the total size of  $C$  plus the number of items in  $K$  which have  $x$ :

$$D(x|C|K) = \frac{\sum_{j \in C} j_x}{|C| + \sum_{j \in K} j_x} \quad (1)$$

In this equation if an attribute  $x$  occurs in all items in  $C$ , but no items in  $C$ ’s contrast set, then  $x$  is fully diagnostic for  $C$  ( $D(x|C|K) = 1$ ). Such an attribute is a perfect guide

to membership of  $C$ : a new item with that attribute is most likely a member of  $C$ . An attribute that does not occur in all members of  $C$ , or which occurs in some members of  $C$ 's contrast set, will be less diagnostic for the category. Such an attribute is a poorer guide to membership of  $C$ : an item with that attribute is less likely to be a category member.

### A logic for evidence

Diagnostic attributes give evidence for an item's classification in a category. Items usually contain a number of different attributes, however, which may be more or less diagnostic for the category in question, or diagnostic for other categories. The diagnostic evidence model uses the following very simple continuous-valued logic to combine the diagnosticity of multiple attributes:

$$\text{NOT } A = 1 - A \quad (4)$$

$$A \text{ AND } B = AB \quad (5)$$

$$A \text{ OR } B = 1 - (1 - A)(1 - B) \quad (6)$$

These equations are simply the probabilistic operations *AND*, *OR*, and *NOT* for samples of independent variables (the probability of *NOT A* is 1 minus the probability of *A*, the probability of *A AND B* is the product of the probabilities of *A* and *B*, when *A* and *B* are independent, and so on).

### Combining Attribute Diagnosticities

To compute an item's overall evidence for membership in a category, the diagnosticity of the item's attributes are combined using the equation for *OR* (Equation 6). An item  $i$  with a set of attributes  $x_1, x_2, x_3$  will be a member of category  $C$  if  $x_1 \text{ OR } x_2 \text{ OR } x_3$  is diagnostic for  $C$ . This is formalised in Equation 7. Let  $A$  be the set of attributes of item  $i$  and let  $D(x|C|K)$  be the diagnosticity of attribute  $x$  for  $C$ . Then  $E(i|C|K)$ , the overall evidence for classifying item  $i$  as a member of  $C$ , is

$$E(i|C|K) = 1 - \prod_{x \in A} (1 - D(x|C|K)) \quad (7)$$

If an attribute  $x$  strictly defines a category  $C$  (occurs in all members of  $C$  and never occurs outside  $C$ ), then  $x$  is perfectly diagnostic of  $C$  ( $D(x|C|K) = 1$ ). If any item  $i$  possesses attribute  $x$ , then by Equation 7  $E(i|C|K)$  will be 1, and the item  $i$  will definitely be a member of category  $C$ . In categories which have no single defining attribute but rather a range of attributes of medium diagnosticity, Equation 7 combines evidence from different attributes in computing evidence for category membership: the more diagnostic attributes the item has, the higher its degree of membership will be.

### Diagnostic evidence in Conjunctions

In the diagnostic evidence model, an item will be a member of a conjunction if it gives evidence for membership in each constituent category in that conjunction: if it has some attributes diagnostic for one constituent category, and other attributes diagnostic for the other. In computing an item's membership in a

conjunction, the model uses the equation for *AND* to combine the item's evidence for membership in each constituent category of that conjunction. An item  $i$  will be classified as member of a conjunction  $C_1 \dots C_N$  if it gives evidence for membership in  $C_1$  *AND* evidence for membership in  $C_2$  *AND* evidence for membership in  $C_3$  and so on. Formally,  $E(i|C_1 \dots C_N | K_{1 \dots N})$ , the evidence for classifying  $i$  as a member of  $C_1 \dots C_N$ , is

$$E(i|C_1 \dots C_N | K_{1 \dots N}) = \prod_{n=1 \dots N} E(i|C_n | K_{1 \dots N}) \quad (8)$$

where the contrast set for the conjunction,  $K_{1 \dots N}$ , is the set of items that are not members of *any* category  $C_1 \dots C_N$ . In computing the evidence for an item  $i$ 's membership in each constituent category of a conjunction (r.h.s. in Equation 8), the contrast set for the conjunction is used. In computing membership in those categories occurring singly, their single contrast sets would be used. This difference in contrast set between single categories and conjunctions is central to the diagnostic evidence model's account for the overextension of conjunctive categories. The next section describes this account in more detail.

### Overextension: Changes in Negative Evidence

In classifying an item in a single category we have one source of positive evidence for membership in the category (the occurrence of that item's attributes in previously-seen members of that category), and one source of negative evidence (the occurrence of the item's attributes in non-members of that category). In classifying an item in a category conjunction, however, the situation is more complicated. To classify an item in a conjunction  $C_1 \& C_2$ , we need two sources of positive evidence for membership: positive evidence for membership in  $C_1$  (which we get if some of that item's attributes occur frequently in stored members of category  $C_1$ ), and positive evidence for membership in  $C_2$  (which we get if some attributes of the item occur frequently in stored members of  $C_2$ ).

We also need to consider negative evidence for membership when classifying an item in a conjunction. Where does negative evidence for an item's membership in a conjunction  $C_1 \& C_2$  come from? It could not come from the item having attributes that occur in stored members of  $C_1$ : as we saw, those occurrences already give positive evidence for membership in the conjunction, and it is a logical necessity that the same thing cannot be both evidence for and evidence against a given conclusion. Similarly, negative evidence could not come from item having attributes that occur in stored members of  $C_2$ : those occurrences already give positive evidence for membership in the conjunction, and the same reasoning applies. So the only source of negative evidence for membership in  $C_1 \& C_2$  must come from previously-seen items that are neither members of  $C_1$  nor members of  $C_2$ : from items that are outside both categories. This is why, in Equation 7, the contrast set for a conjunction consists of all items that are not members of *any* of the constituent categories of the conjunction: logically, only the occurrence of attributes in

**Table 1.** An example set of patient descriptions, with patients classified as having one of 3 different diseases.

Patient Number	Symptoms			Disease Category/ conjunctions
	Eyes	Skin	Muscle	
1	Puffy	Flaking	Strained	<i>Disease A</i>
2	Sunken	Flaking	Knotty	<i>Disease A</i>
3	Sunken	Pallid	Knotty	<i>Disease A</i>
4	Puffy	Sweaty	Knotty	<i>Disease A</i>
5	Puffy	Flaking	Limp	<i>Diseases A&amp;B</i>
6	Puffy	Blotchy	Twitchy	<i>Diseases A&amp;B</i>
7	Red	Flaking	Knotty	<i>Disease B</i>
8	Cloudy	Blotchy	Twitchy	<i>Disease B</i>
9	Red	Blotchy	Twitchy	<i>Disease B</i>
10	Red	Jaundiced	Knotty	<i>Disease B</i>
11	Red	Pallid	Twitchy	<i>Disease B</i>
12	Red	Jaundiced	Weak	<i>Disease C</i>
13	Sunken	Jaundiced	Twitchy	<i>Disease C</i>
14	Red	Flaking	Weak	<i>Disease C</i>
15	Sunken	Flaking	Twitchy	<i>Disease C</i>
16	Sunken	Jaundiced	Weak	<i>Disease C</i>
17	Cloudy	Jaundiced	Twitchy	<i>Disease C</i>

those items can count as evidence against membership in the conjunction. For a single category the contrast set is different: it consists of all stored items that are not members of that single category.

This difference between the contrast sets for single categories and for conjunctions has implications for the classification of items in single categories and in conjunctions. Because of this difference in contrast sets, an item can have negative evidence for membership in a single category (if the item's attribute occur frequently in the contrast set for that single category), but the same item may not have negative evidence for membership in a conjunction containing that category (if the item's attributes do not occur in the contrast set for that conjunction). In this situation the item could be classified as a non-member of the single category (because of negative evidence for membership in that category), but could still be classified as a member of the conjunction, producing a pattern of overextension. In the diagnostic evidence model, a change in negative evidence is reflected in a change in attribute diagnosticity for conjunctions and single categories: this is the source of overextension.

### The source of overextension: an example

I will describe these changes in attribute diagnosticity, and how they cause overextension, using as an example a set of patients suffering different diseases or disease conjunctions. Table 1 shows 17 stored members of three disease categories, *A*, *B* and *C*. These stored members (patient descriptions) are identified by sets of symptoms (attributes), with different symptoms being associated with different diseases. Consider the diagnosticity of the attribute <eyes:sunken> for *Disease C*. <eyes:sunken> occurs in 3 of the 6 members of *Disease C* in Table 1, and occurs 2 times in the contrast set  $K_C$  (the set of items which

**Table 2.** Classification of a patient description in single categories *Disease A* and *Disease C* and conjunction *A&C*.

Evidence for membership in	Patient symptom diagnosticity		
	EYES Sunken	SKIN Flaking	MUSCLES Weak
<i>Disease A</i> singly :	<b>0.56</b>	0.20	0.44
<i>Disease C</i> singly:	<b>0.75</b>	0.38	0.20
<i>Diseases A&amp;C</i> :			
constituent <i>A</i>	<b>0.69</b>	0.29	0.57
constituent <i>C</i>	<b>0.83</b>	0.50	0.33
<i>A&amp;C</i> overall:	<b>0.58</b>		

are not members of category *C*). The diagnosticity of <eyes:sunken> for *Disease C* is

$$D(<eyes:sunken>|C|K_C) = \frac{3}{6+2} = 0.375 \quad (8)$$

This attribute has a relatively low diagnosticity for the single category *C*. Now consider the attribute <eyes:sunken> in context of the disease conjunction *Diseases A&C*.  $K_{A&C}$ , the contrast set for the conjunction *A&C*, consists of items that are members neither of single category *A* nor of single category *C* (items 7 to 11 only). <eyes:sunken> does not occur in any items in  $K_{A&C}$ . The diagnosticity of <eyes:sunken> for *Disease C* relative to the contrast set  $K_{A&C}$  is thus

$$D(<eyes:sunken>|C|K_{A&C}) = \frac{3}{6+0} = 0.5 \quad (9)$$

The attribute <eyes:sunken> is thus more diagnostic for the conjunction *A&C* than for the single category *C* alone.

A similar pattern of diagnosticity change can occur for other attributes, and can lead to an item being given a higher score as a member of a conjunctive category than as a member of one of the constituents of that conjunction (a pattern of overextension). Table 2 illustrates this by showing the computed membership for a patient with the symptoms <found:house>, <eyes:sunken> <skin:flaking> and <muscles:weak>, in the single categories *Disease A* and *Disease C* and in the conjunction *Diseases A&C*. The diagnosticity of the item's attributes for single categories and for constituents of the conjunction (as computed from the attribute distributions in Table 1) are listed in columns under those attributes. The item's membership scores in the single categories and the constituents of the conjunction are computed from those diagnosticities (shown in bold, to the left of those diagnosticities). At the bottom of Table 2 is the item's overall membership score in the conjunction (obtained by multiplying its constituent membership scores). As Table 2 shows, the diagnostic evidence model gives this item a higher overall membership score in the disease conjunction *A&C* (degree of membership 0.58) than in the single disease category *A* (membership 0.56). The model thus predicts overextension for this item in the conjunction *A&C*.

The next section describes an experiment testing the diagnostic evidence model of category conjunction and overextension. This experiment uses as its materials the same disease categories as shown in Table 1.

### Experiment

This experiment looked at overextension in conjunctions of imaginary disease categories (Diseases *A*, *B*, and *C*). In the experiment every participant was given 17 patient descriptions (17 training item) and used these to learn the *A*, *B* and *C* categories. Participants were then given 5 new patient descriptions (5 test items). Participants were asked, for disease categories *A*, *B* and *C* and disease conjunctions *A&B*, *A&C*, and *B&C*, to indicate for each test item whether they thought that test item was a member of that category or conjunction. The frequency of overextension in participants' responses was obtained by counting the number of times a participant classified an item as a member of a conjunction but as a non-member of one or other of the constituent categories of that conjunction.

### Method

**Participants.** 17 Dublin City University students took part.

**Materials.** Training items were presented on 17 printed patient-description cards, each describing a patient with three symptoms. Each patient was identified as suffering from either the disease categories disease *A*, disease *B*, or disease *C*, or suffering from the conjunction of diseases *A&B*. Every participant's set of patient-description cards had the same structure as the set of patient descriptions shown in Table 1. However, the names of symptoms and symptom dimensions were changed for each participant. Every participant saw a different set of patient descriptions, but all descriptions had the same structure.

In the test phase of the experiment, participants were given the 5 new patient descriptions (5 test items) shown in Table 3. Participants were asked to classify these test items in disease categories *A*, *B* and *C* and disease conjunctions *A&B*, *A&C*, and *B&C*. Analysis will focus on how participants classify these test items as members of disease conjunctions and how they classify them as members of the constituent categories of those conjunctions.

**Procedure.** In the first part of the training phase participants spent 15 minutes learning to identify diseases by studying the 17 patient-description cards they had been given. Participants were told to examine the symptoms of patients and to learn to identify a patient's disease from their symptoms. Participants were then given a learning assessment, where they were shown 17 patient descriptions with the same symptoms as those they had learned, but sometimes with incorrect diagnoses. Participants had to identify whether the disease diagnosis given for a patient description was correct or incorrect, by marking a point on a -10 to +10 scale. Questions were presented in random order. If a participant got a question wrong, they were

**Table 3.** Test items to be classified in single disease categories and disease conjunctions in the experiment.

Test Item	Symptoms		
	Eyes	Skin	Muscle
1	Puffy	Jaundiced	Weak
2	Sunken	Flaking	Weak
3	Red	Jaundiced	Twitchy
4	Red	Blotchy	Weak
5	Puffy	Blotchy	Knotty

shown the correct answer, and were later asked that question again.

The test phase of the experiment examined participants' ability to classify the set of 5 new patient descriptions (the test items in Table 3) in the disease categories that had been learned in the training phase of the experiment, and in conjunctions of those categories. This test was run using the same procedure as the learning assessment. Each test-item patient description was presented 6 times, each time paired with a different single-disease or combined-disease diagnosis. At each presentation, participants had to indicate whether the diagnosis shown for that patient description was correct or incorrect by marking a point on a -10 to +10 scale. The names of symptoms and symptom dimensions in test items were different for each participant, and followed the names used in that participant's training items. Questions were presented in random order, subject to the requirement that the same test item was never presented twice in a row. Questions about single-disease diagnoses were equally likely to come before or after those about conjunctions.

### Results

The average number of incorrect responses in the learning assessment part of the training phase was 1.6 ( $SD = 0.76$ ). Participants had little difficulty learning to identify the diseases. To examine the reliability of classification, participants were split into even- and odd-numbered groups. There was a significant Spearman-Brown split-half correlation between the average responses to test-phase items in the two groups ( $r = 0.86, p < .001$ ).

**Applying the model to single-category classification.** To apply the diagnostic evidence model to the classification of test items in the single categories in the experiment, the equations described earlier were used to compute the classification score for each of the test items in Table 3 in the single disease categories *A*, *B* and *C*. The diagnosticity of each test item's attributes for each category was computed from the distribution of those attributes (symptoms) in the training items (as in Table 1). Table 4 shows the observed proportion of participants in the experiment classifying each new patient description (each test item) as a member of each disease category, with the classification scores for those items produced by the model's computations. There was a strong correlation between the model's classification scores and the

**Table 4.** Proportion of participants classifying the 5 test items as members of single categories in the experiment.

Test Item	Category	Proportion classifying item as category member	Model's classification score
1	A	0.23	0.57
1	B	0.15	0.20
1	C	0.73	0.79
2	A	0.12	0.56
2	B	0.15	0.09
2	C	0.81	0.75
3	A	0.04	0.15
3	B	0.77	0.75
3	C	0.62	0.76
4	A	0.00	0.19
4	B	0.50	0.75
4	C	0.73	0.60
5	A	0.77	0.81
5	B	0.62	0.65
5	C	0.08	0.00

classification proportions in the experiment for the single categories ( $r=.86, p <.001, \%var=.74$ ). No free parameters were used in fitting the model to the single category data. (The model's classification scores were also a good fit to participant's average classification scores given to items in the experiment:  $r=.91, <.001, \%var=.82$ ).

**Applying the model to conjunctive classification.** To apply the diagnostic evidence model to the conjunctive category classifications in the experiment, the equations described earlier were used to compute the classification score for each of the test items in Table 2 in the three conjunctions  $A&C$ ,  $A&B$  and  $B&C$ . Table 5 shows the observed proportion of participants in the experiment classifying each test item as a member of each disease conjunction, with the model's classification scores for those items. Again, there was a strong correlation between the model's classification scores and the classification proportions for the disease conjunctions ( $r=.84, p <.001, \%var=.71$ ). No free parameters were used in fitting the model to the conjunctive data. (The model's classification scores were also a good fit to participant's average classification scores given to items in the experiment:  $r=.85, <.001, \%var=.73$ ).

**Overextension in the experiment.** Overextension occurs when an item is classified as a member of a category conjunction, but is classified as a non-member of one of the constituent categories of that conjunction presented singly. For example, if a participant rated a patient with the symptoms <eyes:sunken, skin:flaking, muscles: weak> as definitely suffering from disease conjunction  $A&C$ , but the same participant rated the same patient as not suffering from the single disease category  $A$ , that would be a case of overextension. The occurrence of overextension in the experiment could not be analysed by comparing the proportion of participants who classified an item in a given conjunction with the proportion who classified that item in

**Table 5.** Proportion of participants classifying the 5 test items as members of conjunctions in the experiment.

Test Item	Conjunction	Proportion classifying item as conjunction member	Model's classification score
1	A&B	0.12	0.14
1	A&C	0.58	0.45
1	B&C	0.31	0.22
2	A&B	0.12	0.07
2	A&C	0.65	0.58
2	B&C	0.31	0.08
3	A&B	0.31	0.15
3	A&C	0.04	0.15
3	B&C	0.50	0.81
4	A&B	0.04	0.18
4	A&C	0.08	0.12
4	B&C	0.77	0.56
5	A&B	0.77	0.61
5	A&C	0.04	0.00
5	B&C	0.15	0.00

the constituent categories of that conjunction. Instead, overextension must be examined at an individual participant level. For a given item and a given conjunction, each participant classified that item as a member or non-member of the first constituent category in the conjunction, as a member or non-member of the second constituent category in the conjunction, and as a member or non-member of the conjunction itself. These three classifications constitute that participant's response triple for that item in that conjunction. In the experiment there were 216 response triples in total. Of these, more than one quarter (62 out of 216, or 29%) response triples represented cases of overextension (cases where a participant classified an item as a non-member of one or other constituent category of a conjunction, but classified that item as a member of a conjunction as a whole).

**Most frequent overextension responses.** These overextension responses were not distributed evenly across items and conjunctions in the experiment. Overextension occurred most frequently for test item 2 (symptoms <eyes:sunken, skin:flaking, muscles: weak>). Nearly half of the participants (12 out of 26, or 46%) classified test item 2 as a non-member of single disease category  $A$ , a member of single disease category  $C$ , and a member of disease conjunction  $A&C$ . The next most frequent case of overextension occurred for test item 4 (<eyes:red, skin:blotchy, muscles: weak>). Eight out of 26 participants (30%) classified test item 4 as a non-member of single category  $B$ , a member of single category  $C$ , and a member of conjunction  $B&C$ . All remaining cases of overextension were produced by 4 or less participants.

To examine the diagnostic evidence model's predictions about overextension for in the experiment, the model's classification scores for items in single categories (from Table 4) were compared with its classification scores for those items in the conjunctions (from Table 5). Cases

**Table 6.** Cases where model gives a higher classification score for an item in a conjunction than in a constituent category of that conjunction.

Test item	Model's classification scores in		
2	Category <i>A</i>	Category <i>C</i>	Conjunction <i>A&amp;C</i>
	0.56	0.75	0.58
3	Category <i>B</i>	Category <i>C</i>	Conjunction <i>B&amp;C</i>
	0.75	0.76	0.81
1	Category <i>B</i>	Category <i>C</i>	Conjunction <i>B&amp;C</i>
	0.20	0.79	0.22

where the model produced a higher classification score for an item in a conjunction than it produced for that item in one or other constituent category of that conjunction were taken as predictions of overextension. There were only three such cases (see Table 6). Table 6 shows that, for test item 2, the model produced a lower classification score for that item in single category *A* (a score of 0.56), and a higher score for classification in single category *C* (score of 0.75) and in conjunction *A&C* (score of 0.58). The model thus expected that test item 2 would be classified as a non-member of single category *A* and as a member of single category *C* and conjunction *A&C*. Recall that in the experiment nearly half of the participants classified test item 2 as a non-member of category *A*, a member of category *C*, and a member of conjunction *A&C*. The model was able to predict the most frequent occurrence of overextension in the experiment.

While the first case in Table 6 corresponds to the case in which overextension occurred most frequently in the experiment, the other two cases in Table 6 (test item 3 and test item 1) do not. Examining the classification scores for these test items in Table 6 shows why. For test item 3, the model produced a classification score of 0.75 in single category *B*, a score of 0.76 in single category *C*, and a score of 0.81 in conjunction *B&C* (higher than both constituent classification scores). If overextension occurred whenever an item has a higher classification score in a conjunction than in a constituent category, we would expect overextension for test item 3. However, these classification scores are all very high, for both single categories and conjunction (compare with the other classification scores in Tables 4 and 5). The model would therefore expect this item to be classified as a member of both the single categories and of the conjunction: an overextension response would not occur.

A similar point can be made for test item 1, the final item in Table 6. For test item 1, the model produced a classification score of 0.20 in single category *B*, a score of 0.79 in single category *C*, and a score of 0.22 in conjunction *B&C*. Again, if overextension occurred whenever an item has a higher classification score in a conjunction than in a constituent category, we would expect overextension for test item 1. However, the classification scores for this item in single category *B* and for conjunction *B&C* are both very low. The model would

therefore expect this item to be classified as a non-member of both single category *B* and conjunction *B&C*. Because the item is classified as a non-member of conjunction *B&C*, overextension would not occur.

## Conclusion

The diagnostic evidence model, then, represents a step towards a computational account for overextension in conjunctive categories. The model gave a good fit to people's classification of items in single categories and in conjunctions in the experiment described above, and was able to correctly select the item, category and conjunction for which overextension occurred most frequently in that experiment. While overextension as seen in the experiment seems to be an example of people reasoning illogically about categorization and conjunction, in the diagnostic evidence model overextension is a perfectly logical consequence of the need to use both positive and negative evidence in classification. The model's general approach to overextension may be useful for AI models of classification which aim to reason about conjunctive categories in a way that mirrors human reasoning about such conjunctions.

While the model gives a good fit to the results of the current experiment, work is needed to develop the model into a complete account of overextension in conjunctive categories. Currently missing from the model is a way to consider the absolute values of single and conjunctive classification scores when predicting overextension, as well as the relative differences between those scores. Future work will extend the model in this direction.

## References

- Costello, F. J. 2000. An exemplar model of classification in single and combined categories. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*.
- Costello, F. J. 2001. Testing a computational model of categorisation and category combination: Identifying disease categories and new disease combinations. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*.
- Costello, F. J., & Keane, M. T. 2000. Efficient creativity: Constraint guided conceptual combination. *Cognitive Science*, 242.
- Chater, N., Lyon, K., & Myers, T. 1990. Why are conjunctive categories overextended? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 163, 497-508.
- Hampton, J. A. 1988. Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 55-71.
- Storms, G., De Boeck, P., Van Mechelen, I. & Ruts, W. 1998. Not guppies, nor goldfish, but tumble dryers, Noriega, Jesse Jackson, panties, car crashes, bird books, and Stevie Wonder. *Memory and Cognition*, 26, 143-145.