A Model of Natural Category Structure and its Behavioral Implications[1]

Jane Silber
Management of Technology
Vanderbilt University

Douglas Fisher
Department of Computer Science
Vanderbilt University

Abstract: Fisher (1988) uses the COBWEB concept formation system to illustrate a computational unification of basic level and typicality effects. The model relies on probabilistic, distributed concept representations, and appropriate interaction between cue and category validity. We review this work and report a new account of the *fan effect*. This extension requires an additional assumption of parallel processing, but otherwise is explained by precisely the same mechanisms as basic level and typicality phenomena.

INTRODUCTION

Cognitive modeling fits general computational mechanisms to the constraints of psychological data. The problem of determining an initial starting point for cognitive modeling has been implicitly addressed by several authors. Anderson (in press) suggests a *rational analysis*, whereby a general class of behaviors (e.g., concept formation) are associated with a performance function to be optimized. The guiding assumption is that natural organisms are rational, albeit resource-bounded decision makers.

This paper traces the development of the COBWEB concept formation system (Fisher, 1987) from rational analyses by Gluck and Corter (1985), Kolodner (1983), and Lebowitz (1982). Gluck and Corter provide insights on the absolute quality of conceptual knowledge in their work on human basic level effects. Kolodner's CYRUS and Lebowitz's UNIMEM provide general mechanisms of indexing and classification that we engineer to fit the constraints of basic level effects. In Fisher (1988) the consistency of the resultant model is verified with respect to basic level effects. However, the model also accounts for typicality effects, which were not the focus of engineering. In fact, the model unifies these effects and suggests heretofore unexplored interactions between basic level and typicality phenomena. This paper extends the phenomenological basis of the model by accounting for the *fan effect* (Anderson, 1976). The extensions required for this account are natural, do not adversely affect earlier behavioral accounts, and suggest ways to improve the robustness of COBWEB's underlying learning mechanisms.

BASIC LEVEL EFFECTS AND RATIONAL CONCEPT FORMATION

Substantial experimental evidence suggests that there is a *basic* or preferred level of human classification (Rosch, Mervis, Gray, Johnson, and Boyes-Braem, 1976; Jolicoeur,

[1]Requests for reprints should be sent to Douglas Fisher.

Gluck, and Kosslyn, 1984). For example, when a subject is shown a picture of a collie and asked to name it, the response will typically be *dog*, not *collie*, *mammal*, or *animal*. Similarly, when asked to confirm that a pictured collie is a *collie*, *dog*, *mammal* and *animal*, subjects will respond more quickly for *dog* than the other categories. These tasks indicate that for a hierarchy containing {collie, dog, mammal, animal}, *dog* is the basic level concept.

Gluck and Corter (1985) formulated *category utility*, which presumes that the basic level maximizes 'predictive ability'. For example, very few correct predictions can be made about an arbitrary *animal*, but those that can be made (e.g., animate) apply to many objects. In contrast, knowing something is a *robin* assures many predictions, but they apply to much fewer objects. The basic level concept (e.g., *bird*) is where a tradeoff between the *expected* number of correct predictions (e.g., has-feathers, beaks, flies) and the proportion of the environment to which the predictions apply, $P(N_k)E(\#$ correct predictions$|N_k)$, is maximized. If $P(A_i = V_{ij}|N_k)$ is the probability that an attribute value will be predicted and this prediction is correct with the same probability then this measure can be further formalized as:

$$P(N_k)\sum_i \sum_j P(A_i = V_{ij}|N_k)^2. \qquad (1)$$

Category utility correctly predicts the basic level (as behaviorally identified by human subjects) in two experimental studies (Hoffman and Ziessler, 1983; Murphy and Smith, 1982).

Gluck and Corter's derivation of category utility is motivated by the same rational arguments made by Anderson (in press): good classes are those that maximize correct predictions that can be made about class members. Anderson develops a Bayesian heuristic function to guide concept formation. In contrast, Fisher's (1987) COBWEB uses category utility to guide the incremental formation of classification trees (Kolodner, 1983; Lebowitz, 1982). Fisher (1988) demonstrates that with an appropriate indexing scheme, COBWEB consistently classifies observations at the same intermediate or basic-level classes as human subjects (Hoffman and Ziessler, 1983; Murphy and Smith, 1982).

The indexing strategy is developed from category utility. In particular, (1) can be rewritten (using Bayes Rule) as:

$$\sum_i \sum_j P(A_i = V_{ij})P(A_i = V_{ij}|N_k)P(N_k|A_i = V_{ij}). \qquad (2)$$

Thus, category utility can be viewed as maximizing a weighted (by $P(A_i = V_{ij})$) tradeoff of *cue validity* (i.e., reflected in $P(N_k|A_i = V_{ij})$) and *category validity* (i.e., reflected in $P(A_i = V_{ij}|N_k)$). Indexing can be viewed as 'compiling' this similarity assessment process. Individual attribute value indices are weighted by $P(N_k|A_i = V_{ij})$ and are directed at nodes, N_k, that maximize $P(A_i = V_{ij}|N_k)P(N_k|A_i = V_{ij})$ (i.e., the *collocation* (Jones, 1983)) of the value with respect to ancestors and descendents of N_k. $P(A_i = V_{ij}|N_k)$'s are stored at nodes. Figure 1 illustrates that this strategy results in

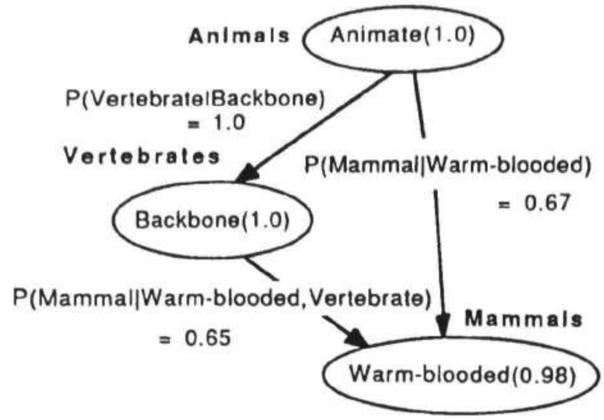

Figure 1: Opportunistic index placement (from Fisher, 1988).

opportunistic indexing [2] that may jump levels. An object is initially classified at that node, N_k, that maximizes the *total cue validity* (Rosch, 1978):

$$\sum_i P(N_k|A_i = V_{ij}), \qquad (3)$$

over the attribute values of the object that are used for indexing. Notice that because category *validity* helps determine index placement it impacts object classification, although it is not explicitly considered at classification time.

TYPICALITY EFFECTS

Importantly, COBWEB does not only account for basic level effects – the phenomena for which it was engineered – but the indexing/classification mechanisms also account for a second influential class of phenomena known as *typicality* effects (Mervis and Rosch, 1981; Smith and Medin, 1981; Rosch, 1978). Psychological studies indicate that some members of a class are treated preferentially or as more typical of a class. For example, in a target recognition task a *robin* will be recognized as a *bird* more quickly than will a *chicken*. In particular, Rosch and Mervis (1975) demonstrate that object typicality increases with the number of features shared with other objects of the same class and varies inversely with the number of features shared with members of contrasting classes.

COBWEB's indexing scheme accounts for typicality effects found by Rosch and Mervis (1975). These studies used letter strings like those of Figure 2a that were arranged into categories A and B and taught to subjects. Subjects were then asked to verify category membership of letter strings of A. Subjects consistently verified membership more quickly for those strings of category A that shared many symbols with other strings of A and shared little with members of category B. To account for this data COBWEB clustered over the collective letter strings of A and B. For example, Figure 2b shows a

[2] A term due to Bareiss, Porter, and Weir (1987).

	Letter String	Intra Over-lap	Letter String	Inter Over-lap	*Typicality*
A	JXPHM	low	4KCTG	high	*low*
	QBLFS	"	GKNTJ	"	"
	XPHMQ	med.	4KC6D	med.	*med.*
	MQBLF	"	HPNSJ	"	"
	PHMQB	high	HPC6B	low	*high*
	HMQBL	"	HPNWD	"	"
B	CTRVG		8SJKT		
	TRVGZ		8SJ3G		
	RVGZK		9UJCG		
	VGZKD		4UZC9		
	GZKDW		4UZRT		
	ZKDWN		MSZR5		
	(1a)		(1b)		

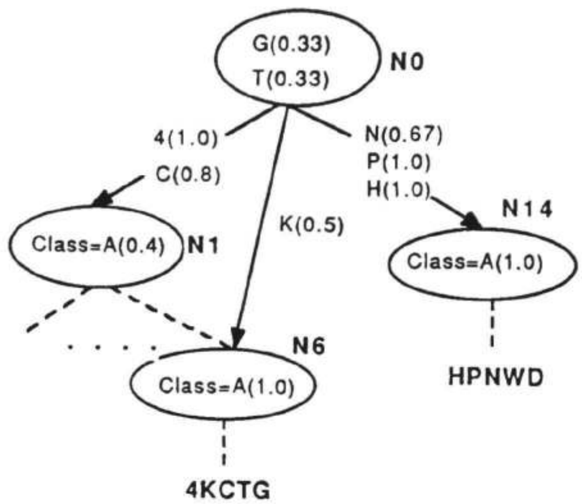

Figure 2: Letter strings and sample COBWEB tree (from Fisher, 1988).

partial tree over the strings of 1*b*. Because some members of category *A* may share more in common with members of *B* than with other members of their own class, class *A* strings are not necessarily localized at a single node. Rather, we assume that a string is recognized (verified) as a category A member by classifying it to a node for which $P(\text{Class}= A|N_k) = 1.0$. Verification time is simulated by the inverse of the total cue validity scores (i.e., 1/total-cue-validity) used to classify the object; we assume that the more an object predicts a node, the faster the object will be classified with respect to it. COBWEB's category-verification time is ordered in precisely the same manner as human subjects, regardless of intra- or inter- category overlap.

On the surface typicality and basic level effects appear to be disparate behaviors. However, Fisher (1988) demonstrates that while concept trees may equate classes with nodes (i.e., a local representation), members of a single class can also be 'distributed' throughout the tree. This enables a unified account of basic level and typicality effects because individual concepts (i.e., the scope of typicality) and concept hierarchies (i.e., the scope of basic level effects) are represented by the same tree-structured representation. This work provides the only computational account of any basic level phenomena that we know of. In addition, the distributed account of typicality effects (with respect to human data found in Rosch and Mervis (1975)) is novel. Finally, the model accounts for known interactions between basic level and typicality effects (Jolicoeur, Gluck, and Kosslyn, 1984) and predicts previously unexplored interactions.

FAN EFFECT

Work since (Fisher, 1988) has accounted for a third phenomena: the *fan effect* (Anderson, 1976). The fan effect has been demonstrated in sentence recognition tasks. Typically, simple sentences that consist of a person and a location are used:

(1-1) The doctor is in the bank. (1-2) The fireman is in the park.

(2-1) The teacher is in the church. (2-2) The teacher is in the park.

The sentences vary in the number of features associated with the subject of the sentences and the location in which the subject appears (e.g., 'teacher' appears in two sentences). The numbers following each sentence indicate the size of the fan: the number of sentences that contain the feature (person – location). After training on selected sentences, recognition experiments are performed; subjects must respond as to whether they have previously observed a sentence (true) or not (false). Recognition time increases with the frequency that a person and location is present in training sentences.

For COBWEB, sentences are encoded as attribute value pairs. The set of objects, each of which contains two attributes (i.e., person and location), is then used to create a concept tree. A test set is a mixture of items that appeared in the original training set ("trues"), and new sentences that have not been seen previously ("falses").

One key processing assumption was added to the basic classification model. Many studies in cognitive psychology have suggested that search through memory proceeds in a parallel fashion. Triggering nodes in memory will cause activation to spread among all related elements, perhaps with different degrees of strength or speed. The assumption of parallel search was added to the COBWEB model. Rather than only examining the path that maximizes total cue validity, all paths indexed by object (sentence) values are explored. The search ends when indices lead no further or when the test item is found in a node. For the "true" statements, COBWEB always locates the test object in a node, thus ending the search. The total simulated time required to reach that node is the resultant recognition time. The search for "false" test objects, on the other hand, will end when all paths have been explored as far as possible. In these cases the limiting factor is the time required to explore the *slowest* path. Our data compares favorably with experimental data, in cases of true (observed) and false sentences and across all feature frequencies. Figure 3 contains a portion of the tree produced by COBWEB when presented with a set of person-location sentences. The dotted lines in the diagram represent the indices that are used to recognize a test probe. In the training set, *doctor* and *church* each appear in only one sentence, while *park* appears in two. When the "false" probe *The doctor is in the church* is presented, COBWEB predicts that the search will simultaneously follow both the *doctor* index and the *church* index, leading from *N0* to *N3* and *N6*. Both of these paths are exhausted with a total time of 1 unit. In contrast, the "false" probe *The doctor is in the park* has a longer response time, because *park* appears in two sentences and has a larger fan. The search resulting from this probe proceeds from *N0* to *N3* along the *doctor* index, requiring 1 unit of time. However, the search simultaneously follows the *park* index from *N0* to *N2*, requiring 1 unit of time, and then from *N2* to *N7*, requiring another unit. Therefore, 2 units of time are required before *N7* is reached and the model can identify the probe as false.

Table 1a shows the mean recognition times for "true" and "false" statements in actual human experiments (Anderson, 1976). In comparison, the (unfitted) reaction times

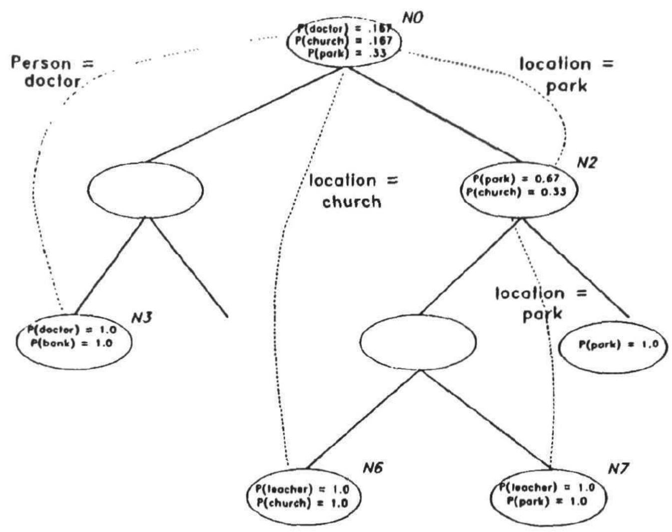

Figure 3: Concept tree for person-location experiment.

Table 1: Fan effect mean reaction times. Mean time for 'true' statements are shown above the mean time for 'false' statements.

(a)

Sentences / location	Sentences/person: 1	2	3
1	1.11	1.17	1.22
	1.20	1.22	1.26
2	1.17	1.20	1.22
	1.25	1.36	1.29
3	1.15	1.23	1.36
	1.26	1.47	1.46

(b)

	Sentences/person: 1	2	3
1	0.50	0.83	0.95
	1.00	1.50	2.25
2	1.47	2.30	2.25
	1.55	1.70	1.82
3	1.65	2.30	2.85
	1.62	1.97	2.07

predicted by the COBWEB model are displayed in Table 1b. COBWEB produced a concept hierarchy from the same training set used in human experiments; the data presented here are averaged over several trials. We expect systematic increase in time as the number of sentences per person and per location increase, comparisons of relative magnitude are most meaningful. In the recognition time tables, there are 36 possible comparisons of relative size (18 each for "trues" and "falses"). Human experiments and COBWEB simulations each resulted in 3 comparisons that are not in the expected direction.

There is great similarity between the COBWEB account of the fan and typicality effects. Typicality studies are generally based on target recognition tasks that require subjects to classify an instance as a member of a category. Instances with high inter-category similarity are associated with longer response times, while high intra-category similarity produces shorter response times. On the other hand, Anderson's (1976) ACT model predicts that instances with a large fan resulting from many associated propositions will have longer response times; ACT accounted well for the human data. This produces an apparent contradiction in that objects with features shared with many other objects (i.e., persons or locations appearing in many sentences) produce longer times in the fan effect, but are apparently more "typical", thereby resulting in shorter times according to the typicality effect. However, further examination of the learning task reveals that these two findings are consistent, and the explanation rests on the distinction between intra- and inter- category similarity. When propositions or sentences are learned in fan effect studies, each is remembered as an individual case, or category. Persons or locations that appear in a large number of sentences correspond to attributes that are common to more than one category, i.e., high inter-category similarity not intra-category similarity. Thus, the direct relationship between fan size and response time closely parallels the relation between typicality and inter-category similarity. The COBWEB model accounts for typicality and fan effects in the precisely the same manner; the fan effect emerges as a special case of typicality effects in which the classes being learned are singletons. Although the original COBWEB typicality studies were conducted without the parallel processing assumption (Fisher, 1988), similar results are obtained when parallelism is incorporated.

CONCLUDING REMARKS

We have extended the scope of behaviors accounted for by COBWEB. By our account, the fan effect is a special case of typicality phenomena. We are extending our research in several directions. First, computer experiments reveal that very early in concept formation our indexing scheme is very sensitive to the ordering of observations. Indexing is easily fooled and led astray. In general, our indexing procedure and tree structure are too inflexible. Early in training desirable classes can fluctuate wildly. Our work with the fan effect suggests that rather than placing (classifying) an object along a single best path, it may be more desirable to place (classify) it along a number of paths. In fact, the category utility indexing scheme is easily extensible to allow this – without the use of arbitrary thresholds that characterize other systems (Kolodner, 1983; Lebowitz, 1982). Classification along multiple paths leads naturally to a directed acyclic graph structure (DAG). A DAG is more robust in that is allows orthogonal classes to develop (e.g., mammal or reptile or bird or ... or fish *versus* carnivore or omnivore or herbivore). Classes that do not prove useful later in training can be pruned out. Thus, a rational analysis (Anderson, in press; Gluck & Corter, 1985) initially led to a model of certain psychological effects, but an inverse process is also valuable: modifications to the cognitive model suggest extensions that are primarily computational improvements.

REFERENCES

Anderson, J. R., *Language, Memory, and Thought*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1976.

Anderson, J. R., "The Place of Cognitive Architectures in a Rational Analysis." In K. Van Lehn (ed.), *Architectures for Intelligence*, (in press).

Bareiss, E.R., Porter, B.W., and Wier, C.C. Protos: an exemplar-based learning apprentice. In *Proceedings of the Fourth International Workshop on Machine Learning*, Morgan Kaufmann, 1987, pp. 12–23.

Fisher, D., "A Computational Account of Basic Level and Typicality Effects," *Proceedings of AAAI-88: The Seventh National Conference on Artificial Intelligence*, Morgan Kaufmann, 1988, pp. 233-238.

Fisher, D., Knowledge Acquisition Via Incremental Conceptual Clustering, *Machine Learning*, 2, 1987, pp. 139-172.

Gluck, M. A. and Corter, J. E., "Information, Uncertainty, and the Utility of Categories," *Proceedings of the Seventh Annual Conference of the Cognitive Science Society*, Irvine, CA, Lawrence Erlbaum Associates, 1985, pp. 283-287.

Hoffman, J., and Ziessler, C., "Objectidentifikation in Kunstlichen Begriffshierarchien," *Zeitscrift fur Psychologie*, 16, 1984, pp. 43-275.

Jolicoeur, P., Gluck, M., and Kosslyn, S., "Pictures and Names: Making the Connection," *Cognitive Psychology*, Vol. 16, 1984, pp. 243-275.

Kolodner, J. L. (1983). Reconstructive memory: A computer model. *Cognitive Science*, 7, 281–328.

Lebowitz, M. (1982). Correcting erroneous generalizations. *Cognition and Brain Theory*, 5, 367–381.

Murphy, G., and Smith, E., "Basic Level Superiority in Picture Categorization," *Journal of Verbal Learning and Verbal Behavior*, Vol. 21, 1982, pp. 1-20.

Rosch, E. and Mervis, C., "Family Resemblances: Studies in the Internal Structure of Categories," *Cognitive Psychology*, Vol. 7, 1975, pp. 573-605.

Rosch, E., Mervis, C., Gray, W., Johnson, D. and Boyes-Braem, P. "Basic Objects in Natural Categories," *Cognitive Psychology*, Vol. 8, 1976, pp. 382-439.

Smith, E. and Medin, D., *Categories and Concepts*, Harvard University Press, Cambridge, MA, 1981.