Facilitation from Clustered Features: Using Correlations in Observational Learning[1]

Dorrit Billman

University of Pennsylvania

Evan Heit

Stanford University

Jennifer Dorfman

University of California, San Diego

In learning categories and rules from observation without external feedback, people must make use of the structure intrinsic to the instances observed. Success in learning complex categories from observation, as in language acquisition, suggests that learners must be equipped with procedures for efficiently using structure in input to guide learning. We propose one way that learners make use of the correlational structure available in input to facilitate observational learning: increase reliance on those features discovered to make good predictions about the values of other features. Such a mechanism predicts two types of facilitation from multiple correlations among input features. This contrasts with the effects of correlated features which have been suggested by models addressing learning with explicit feedback. Three experiments investigated learning the syntactic categories of artifical grammars, without external feedback, and tested for the predicted pattern of facilitation. Subjects did show the predicted facilitation. In addition, a simulation of the learning mechanism investigated the conditions when it would provide most benefit to learning. This research program begins investigation of procedures which might underlie efficient learning of complex, natural categories and rules from observation of examples.

Introduction

How do people learn about complex categories and rules from observation of examples? Adept performance in a variety of domains-- from linguistic to social-- suggests that people do abstract categories that reflect the organization in input. Little is known about learning from observation without benefit of explicit feedback or tutoring since research on learning, both in psychology and in artificial intelligence, has focused on explicit learning tasks with direct feedback provided. The present research explores observational learning without feedback and investigates how people use correlational structure available from examples to learn categories and rules.

Several researchers have suggested that category structure for psychologically coherent, natural catgories is rooted in the correlational structure provided by features in input (Rosch, 1978, Medin, 1983). Discovery of correlations among multiple features may be particularly important for observational learning. Maratsos and Chalkley argue that syntactic categories are acquired by extracting the system of interpredictive relations which hold among members of the same syntactic class (Maratsos & Chalkley, 1980).

Our work begins with these suggestions that correlational structure is important, and proposes how it might be used in learning without feedback. We use experimentation and simulation to test two predictions about the effects of correlational structure and to investigate learning procedures responsible for these effects.

Both predictions claim that availability of multiple, intercorrelated features in input will facilitate learning. We propose two levels of facilitation, one general, and one specific to individual rules about feature covariation. The first prediction would be anticipated by many learning models: with multiple cues people should be more likely to learn at least one of the feature covariation rules, or at least something about the available categories. The odds of discovering at least one rule, or some basis for categorizing, are better when there are multiple regularities to be discovered (Trabasso & Bower, 1968).

The second prediction points to a stronger type of facilitation from multiple covarying cues and a more powerful method of capitalizing on correlational structure. This predicted pattern will be called clustered feature facilitation. It claims that subjects will be more likely to discover a target correlational rule when that rule occurs in a context which provides other rules among the same set of features. For example, applying the principle to syntactic categories would imply that learning the relations between *semantics* and *phrase structure rules* would be facilitated in a system which has, in addition, 1) covariation rules between *morphology* and *semantics* and 2) rules between *morphology* and *phrase structure rules*.

One mechanism which would produce this rule-by-rule facilitation is an attentional learning procedure, called focused sampling (Billman, 1983), which increases attention to predictive features. If a feature is predictive in one covariation rule it would become favored, or selected more often, in testing other rules. When multiple rules hold among a set of features they would provide a mutually reinforcing effect, increasing the prominence of the valuable, predictive set of features. A mechanism of this sort would result in mutual facilitation among correlated cues. This prediction contrasts with the pattern found for learning simple concepts with feedback (Trabasso & Bower, 1968): here, attending to and learning about one cue was independent of learning about a second, covarying cue.

Our research orientation and the prediction of clustered feature facilitation differ from much category learning research in two ways: we focus on learning without external feedback and we argue that successful learning here requires capitalizing on covariation among multiple cues. Specifically, we argue for a learning mechanism which uses covariation among multiple cues to facilitate rule and category learning, rather than one where learning about one cue is independent of the relations among others (Trabasso & Bower, 1968), or where learning about one cue only competes with learning about covarying cues (Zeaman & House, 1963, Lovejoy, 1966). The experiments test for facilitation from multiple covarying features.

Experiments

Rationale

Three experiments tested for general facilitation and for clustered feature facilitation. They investigated learning of syntactic categories in an artificial language. This allowed control of the regularities from which rules might be induced and maximized the chances of tapping a domain where people are naturally successful observational learners. Structures in the grammars were analogous to structures which occur in natural languages. All experiments investigated acquisition of rules distinguishing two relational classes. All experiments presented scenes with descriptive sentences which subjects observed to learn the language. All experiments compared two conditions, Structured and Isolating. All experiments compared learning of 15 college student subjects in each condition. Conditions differed in the correlational structure afforded by input. For example, Experiments 1 & 2 compared learning two relational classes (analogous to verbs and prepositions). Each class participated in a distinct set of phrase structure rules and had distinct semantics--referring to action or to relative position. In the Isolating Conditions (Experiments 1 & 2), semantics and phrase structure role were the only characteristics which correlated with one another or which could serve as a basis for learning the syntactic categories. In the Structured Conditions, additional features such as morphology, agreement rules, and syntactic marker words covaried with semantics and phrase structure rules. The two predictions were tested by comparing learning in Structured and Isolating Conditions. Some tests assessed general facilitation. Others assessed knowledge of just the target rules common to both conditions and tested clustered feature facilitation.

Table 1

Rules and Vocabulary for Structured & Isolating Conditions

Experiments 1 & 2

Examples for the Structured Condition, with English Glosses

Phrase Structure Rules with Examples of Sentence Types

1) $NP_{AG} + PP_{OB} + PP_{LC}$	(DOBOD) (BO SAFAT VULK) (BO LARAN NINK) The DOBO-subj is beside the SAFA-obj, below the LARA-loc.
2) $NP_{AG} + PP_{OB}$	(DOBOD) (BO SAFAT VULK) The DOBO-subj is beside the SAFA-obj.
3) $NP_{AG} + PP_{LC}$	(DOBOD) (BO LARAN NINK) The DOBO-subj is below the LARA-loc.
4) $NP_{AG} + PP_{LOC} + VP$	(DOBOD) (BO LARAN NINK) (PIR SAFAT TOFO) The DOBO-subj is below the LARA-loc and exchanges places with the SAFA-obj.
5) $NP_{AG} + VP$	(DOBOD) (PIR SAFAT TOFO) The DOBO-subj exchanges places with the SAFA-obj.

Relational Vocabulary & Characteristics

N-**Verb**	**Structured**	**Isolating**
Characteristics:	VP uses, actions, verb & marker agree, two syllables, PIR or TEW.	VP uses, actions. No Other Consistent Pattern.
Vocabulary:	*PIR/TEW ... TOFO/TAFA* *PIR/TEW ... BOPO/BAPA* *PIR/TEW ... JOSO/JASA*	*BO ... GORK/GARK* *PIR/TEW ... BOPO/BAPA* *PIR/TEW ... JUS*
N-**Prep**	**Structured**	**Isolating**
Characteristics	PP uses, positions, no agreement, one syllable, BO.	PP uses, positions. No Other Consistent Pattern.
Vocabulary:	*BO ... GIRK* *BO ... NINK* *BO ... VULK*	*PIR/TEW ... NINK* *BO ... TIF* *BO ... VALK/VOLK*

Note: / indicates two forms of a verb or marker. ... indicates a missing shape word.

Experiments 1 & 2

Method

Stimuli

Sentences in "Neptunese" described scenes with triples of interacting objects. Three transitive actions were used, in which an agent shape moves to act on a remote object while a shape near the agent remains inactive. Sentences described the actions and relative positions among the shapes. Words referring to objects were inflected for case.

Both Structured and Isolating Conditions had two syntactic classes of relational words which had different roles in the phrase structure rules. In the Structured Condition multiple features covaried and provided a basis for grouping these words into contrasting categories. In the Isolating Condition only a subset of these features covaried, namely, semantics and distribution in phrase structure rules. In both conditions, semantics of referent and phrase structure rules covaried and distinguished two syntactic categories. Words referring to actions had one set of privileges of occurrence in phrase structure rules and will be called *N*-Verbs (for Neptunese verbs). Words referring to relative positions had contrasting (though overlapping) phrase structure rules and will be called *N*-Preps. Table 1 shows the f phrase structure rules and the five possible sentence types which can describe a scene.

Structured and Isolating Conditions differed with respect to regularities among other features, as summarized in Table 1. In the Structured Condition, additional features covaried with semantics and phrase structure distribution. All *N*-Verbs had two syllables and the vowels in the syllables changed to agree with the subject; e.g. "TOFO" was used when nouns of one (phonologically defined) gender class were used as subject and "TAFA" was used with the other gender. All *N*-Verbs used "PIR" or "TEW" as their syntactic marker at the beginning of the verb phrase, with selection between PIR and TEW again based on subject gender. In contrast, *N*-Preps were monosyllables ending in "K"; they had "I"'s and "U"'s in their stem rather than "O"'s or "A"'s; they did not change form with subject gender; and they had a single, fixed marker "BO".

In the Isolating Condition, each word had fixed phonology, agreement rules, and marker element; indeed, the values of these features here were identical to those used in the Structured Condition. However, there was no systematicity in the manner by which these features were assigned to words within a class; none of these features covaried with semantics or phrase structure distribution. *N*-Verbs had one or two syllables, fixed or changing forms, markers of "BO" or of "PIR"/"TEW"; similarly, various *N*-Preps had all possible values of these features. No pair of these features covaried with another or with semantics and phrase structure rules.

In the Structured Condition, semantics, phrase structure distribution, phonology, agreement, and marker word all covaried. In the Isolating Condition, only semantics and phrase structure distribution covaried.

Procedure

Subjects in either condition learned by watching scenes and reading out loud the accompanying descriptive sentences. The learning phase ran for two hours on each of two days and included some auxillary tasks (e.g. writing sentences). Following exposure to either the grammar of the structured or isolating condition, tests assessed what subjects had learned about the covariation patterns defining the classes. Three tests will be presented here. In all tests subjects were presented with a novel display and asked to judge whether it fit in with the learning sentence or whether it contained any error. Subjects rated sentences on a six-point scale in which one end point indicated the subject was certain the item was correct and fit in with the learning sentences and the other end point indicated certainty about an error.

The Phrase Structure Test assessed any knowledge for distinguishing between classes. Comparison between conditions tests the predicted general facilitation from multiple correlated cues. Subjects judged novel sentences using familar words but presented without any picture. They decided whether the new sentence fit into the language and was correct or whether it did not belong and contained an error. Errors were created by substituting a *N*-Verb into the context where a *N*-Prep was required or a *N*-Prep into a context for *N*-Verbs. Subjects could detect errors if they knew any of several rules, since the incorrectly used word might lead to inconsistent assignments among phonological properties, marker word, and phrase structure role. Subjects in the structured conditions might be able to detect errors based on disruption of any of several relations among predictive features, while subjects in the isolating condition would only be able to detect an error based on misplacement of the particular word in the phrase structure tree.

Two tests assessed clustered feature facilitation: the Projected Use Test and the Semantics Test. Experiment 1 used only the Projected Use Test; Experiment 2 used both. These tests evaluated knowledge of a single component of the grammar, one available to learners in both conditions.

The Projected Use Test assessed knowledge of phrase structure rules, rules which were identical in both conditions. The subject's task was to judge compatibility of a second, projected use of a novel word given an introductory use. Use of novel words means subjects cannot make judgements based on known uses of the word, but must rely on more abstract knowledge of relations among sentence forms. The task was explained by analogy to how we can make inferences about available uses of new words in English. In English, if one hears a new word, "glish", used in "She met a glish boy", one could be fairly certain that it would be grammatically okay to say "That girl is very glish." In the example, correct judgment requires use of implicit knowledge of the uses allowed for English adjectives; in the test, analogous judgments assessed knowledge of contrasting uses allowed for one of the two Neptunese categories. Subjects were told they would be given one correct, introductory use of a new word and then would be asked to judge whether a second use of the new word was also okay.

Testing clustered feature facilitiation means specifically testing knowledge of the relation among target features available to subjects in both conditions. Hence any other features which might be differentially informative for the two conditions must be eliminated. Semantic cues were eliminated by not displaying any picture. Phonological cues were removed by simply indicating the new word by the symbol *NEW*, which appeared in the sentence at the location of the target word, rather than spelling out the new word. Cues from the marker words were removed by indicating its presence by the symbol ////, without showing which marker was used. These conventions were again explained by analogy to inferences we make listening to speech; sometimes we can tell quite a bit about a new word even if we don't catch all the words in the sentence. English analogies, examples using the contrast between Neptunese relational terms versus nouns, and explaining the task back to the experimenter ensured that all subjects understood the task.

Removal of information about phonology, marker words, and agreement rules for the novel test word is critical to a fair test of clustered feature facilitation. The intent is to test whether subjects in the structured condition were more likely than those in the isolating conditions to learn an identical set of target rules (here, the phrase structure rules). While the target rules were identical in both conditions, additional cues covaried with the target rules in the Structured Condition. Clustered feature facilitation claims that the identical rule will be learned faster when it occurs as part of a related system of rules among the same features. A test of clustered feature facilitation requires that information available during learning differs between conditions, but information available at test is identical. Differential performance at test can then be attributed to differential learning about the same cues, not the availability of more cues to use in detecting errors.

The Semantics Test assessed learning the relation between semantics and the set of acceptable uses (phrase structure rules); action words are used in one set of rules, position words in another. This component of the grammar was the same in both Structured and Isolating Conditions. As in the Projected Use Test, the clustered feature prediction is tested by asking subjects to judge novel displays where correct judgments could only be made based on the target rule, common to both conditions. Subjects were told they would see scenes with some new aspect and a descriptive sentence which included a new word describing the new aspect. The new word was indicated by *NEW*. Subjects judged whether the word for the new aspect was used correctly. Subjects could judge correctly if they knew which uses were acceptable for words referring to action and which for words referring to position. Additional information about syllables, agreement, and markers was eliminated.

The Projected Use Test evaluates whether subjects are better able to learn the covariation between syntactic contexts shared by the same word (the phrase structure rules) when additional features covary with the features of the target rule. The Semantics Test evaluates whether subjects are better able to learn the covariation between contexts of use defined by phrase structure rules and semantics when additional features covary at learning. Both test the predicted clustered feature facilitation.

Results

Both experiments found a strong general advantage for learning in a grammar where multiple cues covaried to define syntactic categories. Subjects in the Structured Condition performed better than those in the Isolating Condition on the Phrase Structure Test, in both experiments. Both experiments also provided evidence for clustered feature facilitation. Subjects from the Structured Condition performed better on the Projected Use Test, in both experiments. The Semantics Test consisted of four types of items: correct action, correct position, incorrect action, and incorrect position displays. Structured Condition subjects did better than Isolating Condition subjects on two of the four measures, but did not differ on the summary measure. See Table 2.

Table 2

Scores for Experiments 1 & 2

	STRUCTURED CONDITION 1 mean(s.d.)	ISOLATION CONDITION 2 mean(s.d.)
	Experiment 1	
PHRASE STRUCTURE (48 items)	4.66(.75) *	3.80(.41)
PROJECTED USE (48 items)	4.03(.74) *	3.63(.29)
	Experiment 2	
PHRASE STRUCTURE (48 items)	4.59(.84) *	4.10(.50)
PROJECTED USE (48 items)	4.19(.76) *	3.83(.23)
SEMANTIC (48 items)	4.44(.29)	4.22(.21)
action/correct	5.44(.51) *	4.97(.61)
position/correct	4.17(.73)	4.57(.76)
action/incorrect	4.26(1.00)	4.14(.71)
position/incorrect	3.90(.86) *	3.19(.79)

* Structured Condition better than Isolating Condition 2 ($p<.05$)

Average scores on a 1 to 6 rating scale. Score of 6 means subject was certain every grammatical sentence was correct and every ungrammatical sentence was incorrect.

Experiment 3

Experiment 3 investigated the contrast between two subcategories of verbs. As in Experiments 1 & 2, learning a target rule was compared in Structured and Isolating Conditions. In Experiment 3 the relation between phrase structure rules and lexical form was preserved in both conditions, rather than the relation between the phrase structure rules and semantics, as in Experiments 1 & 2. Phrase structure forms are shown schematically here:

Class 1: $S \rightarrow NP_{AG} + NP_{OB} + (\text{Marker } NP_{LC}\ V_1)$
$S \rightarrow NP_{AG} + (\text{Marker } NP_{LC}\ V_1)$
$S \rightarrow NP_{AG} + NP_{OB} + V_1$
Class 2: $S \rightarrow NP_{AG} + NP_{LC} + (\text{Marker } NP_{OB}\ V_2)$
$S \rightarrow NP_{AG} + (\text{Marker } NP_{LC}\ V_1)$

Method

In the structured condition, Class 1 words had a distinct set of allowable uses or phrase structure rules; they referred to transitive actions; they had unchanging, single-syllable lexical forms which had "I"'s or "U"'s for vowels; and they used the marker "BO". Class 2 words had contrasting phrase structure rules; they referred to indirect causative actions; they had two syllables; they had agreement rules which changed the verb vowels to "O"'s or "A"'s to agree with the gender of the subject; and they used the markers "PIR" or "TEW", selected to agree with the subject's gender as well. In the Isolating Condition, only one of these characteristics covaried with phrase structure rules: agreement. One class had agreement rules which changed the form of the verb and the other class had fixed forms with no agreement rule. Number of syllables, marker word, and semantics were unsystematically assigned; they did not covary with each other, with uses defined by the phrase structure rules, or with fixed or changing phonological form.

Learning and test procedures were exactly analogous to those used in Experiments 1 and 2. The Phrase Structure Test assessed general facilitation in learning any rules relevant to the two syntactic categories (verb subcategories). The Projected Use Test assessed learning the dependencies among different uses (phrase structure rules) to see if learning a target rule was facilitated by the availability of additional covarying features in learning. The analog of the Semantics Test was the Lexical Test. The Lexical Test assessed subjects' knowledge of the covariation rules between lexical form (fixed or varying with agreement) of the word and allowable position in the phrase structure tree. In the Lexical Test, subjects were presented with sentences using a new word. Only the first part of the word was shown, enough to reveal whether or not it agreed with the subject but not enough to tell whether it was one or two syllables. Information about marker and semantics was also deleted. Subjects could be consistently correct only if they knew the relation between position in the sentence and whether or not the word should agree with the subject.

Results

Subjects in the Structured Condition did better than those in the Isolating Condition on all tests, the Phrase Structure Test, the Projected Use Test, and the Lexical Test. Results are summarized in Table 3.

Table 3

Scores for Experiment 3

	STRUCTURED CONDITION 1 MEDIANS	ISOLATION CONDITION 2 MEDIANS
PHRASE STRUCTURE (48 items)	5.90 *	3.90
PROJECTED USE (48 items)	5.52	4.00
Class 1 uses	mean=5.04 **	mean=4.30
Class 2 uses	mean=5.06	mean=4.48
LEXICAL (24 items)	3.96 *	3.46

*Structured Condition median better than Isolating Condition 2 ($p<.05$)
Medians used due to nonnormal distributions.

**Structured Condition mean better than Isolating Condition 2 ($p<.05$)

Average scores on a 1 to 6 rating scale. Score of 6 means subject was certain every grammatical sentence was correct and every ungrammatical sentence was incorrect.

Discussion of Findings

The results from the Phrase Structure tests found general benefit to learning about the system from the availability of multiple correlated features, consistent with several prior findings (Morgan & Newport, 1981, Green, 1979). The benefit was found for two pairs of syntactic classes and when semantics or morphological features covaried with distributionally defined class. In addition, the Projected Use, Semantics, and Lexical Tests suggest that when multiple features covary, learners do capitalize on the availability of one covariation pattern to discover others. Subjects did show clustered feature facilitation.

Possible Learning Mechanism

Focused sampling is part of the internal feedback approach (Billman, 1983) which specifically addresses learning without external feedback. The internal feedback approach predicts clustered feature facilitation by combining generation and testing of predictive rules with attentional learning. The learner projects hypotheses about the expected value of a second (set of) feature(s), given the value of a first. These hypotheses are evaluated by comparing the predicted with the observed value of the projected feature and using this match or mismatch as internally generated feedback. Hence the first level of learning consists of generating and testing conditional rules about predictive relations among features. Focused sampling, the attentional learning procedure, systematically alters the sampling of features used in hypothesis generation and testing. Whenever the prediction of a rule is confirmed, focused sampling increases the salience (or sampling probability) of the features which participate in that rule. Following bad predictions, salience of participant features is reduced. Focused sampling will benefit rule learning when successful participation in one rule is in fact predictive of that feature's participation in other rules. Where input does provide a system of covariation rules among an overlapping set of features, focused sampling will lead to learning the individual rules faster than when input provides only one of the individual rules by itself. Rule learning with focused sampling leads directly to the predicted clustered feature facilitation.

This account of attentional learning contrasts both with most featural and most instance-based views. Typically, models do not propose that the importance of or attention to a feature *increases* in the context of other covarying features. It may not change systematically at all (Medin & Schaffer, 1978, Trabasso & Bower, 1968), or covarying features may compete with each other (Zeaman & House, 1963, Fisher & Zeaman, 1973). Direct comparison to these models is not possible because the task changes when no feedback is available; extending these models to address learning without feedback might indeed prompt modifications in how the learner uses correlational structure. While the finding of clustered feature facilitation might be predicted by other types of models, it contrasts with a broad range of alternative types.

Simulation

Rationale

The exact effects of focused sampling are not transparent from an informal statement, so we turned to simulation. Simulation, unlike experiments with people, allows addition and deletion of a component at will; hence the contribution of that component can be directly assessed. The simulation compares two versions of a learning mechanism, with and without focused sampling. We adopted a design perspective and varied the learning problems to ask whether and when focused sampling would provide substantial facilitation (Billman & Heit, 1987).

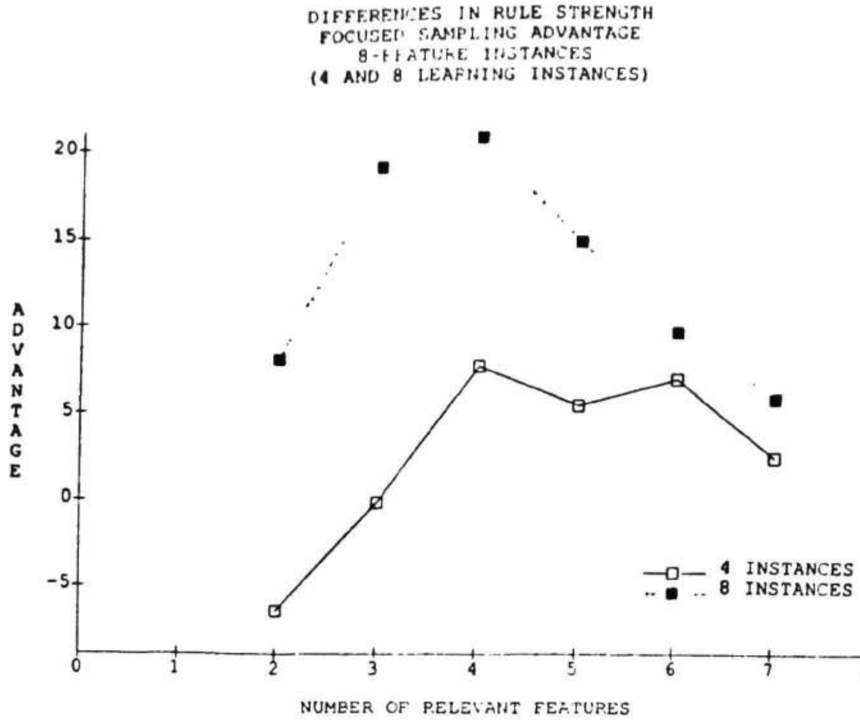
Input for the simulation was abstract, schematically specified examples. Each example consisted of a set of features; each feature assumed one of four values. With focused sampling feature salience, or importance, changes with learning. Here each learning cycle consisted of these steps: 1) one predictor feature is sampled and, given its value, a prediction made about the value of a second feature; 2) the prediction is tested against the value specified in the input example; and 3) salience of the features as well as confidence in the prediction are both modified. Rule strength is increased or decreased depending on success. The saliences of features in the rule are also increased or decreased, changing their probablity of sampling; salience across features is normalized after each change, producing indirect, compensatory change too. The relation among features can be both supportive and competitive: supportive because of convergence on a common set of rules via focused sampling, competitive because limited attention is still allocated among the features. Without focused sampling the learning cycle does not include modifying feature salience, but is otherwise identical.

Findings

A series of experiments with varying learning stimuli were run to compare learning with and without focused sampling. Input sets varied in number of instances, number of features, and most importantly in relative number of correlated and uncorrelated features. Across a wide range of learning stimuli, focused sampling did produce benefit. Amount of benefit was assessed by comparing difference in learning with and without focused sampling at a criterion defined to match amount of exposure in the two conditions. The pattern of factors resulting in more or less benefit can be summarized in four points.

First, where comparable conditions were run varying in number of instances (4, 8, and 16), focused sampling benefit was greater for the larger number of instances.

Second, benefit was greater when input instances had more features, particularly for the first level of increase from very few (e.g. 4) to somewhat more (e.g. 6); effect of adding more features leveled off with larger numbers (e.g. 12).



Third, relative number of correlated and unpredictive features had major effects on benefit. The effects of adding more relevant or more irrelevant features interact with each other and with total number of features; only the major points are summarized here. For example, when the number of total features is fixed at 8, benefit increases as correlated features increase from 2 to 4, but benefit decreases as still more of the total are made correlated, as shown in Figure 1. Benefit increases as more features of a fixed total are made relevant. In addition, benefit increases as new relevant features are added. However, this pattern holds only for the initial increase (e.g. maximally for the increase from 2 to 3 relevant features) and attenuates or decreases with increasing additions. Benefit also increases as more *irrelevant* features are added, when irrelevant features initially are few in number. In sum, benefit is greatest when there are multiple correlated features but there are still a significant proportion of irrelevant features (from which the relevant features can steal attention). Thus, this simulation produced clustered feature facilitation over a wide range of conditions, but only where a significant number of "unsystematic" features were also present.

Finally, we found one set of conditions where focused sampling *hurt* learning the relation between the two correlated features, the target rule. This occurred when only four instances were in the learning set, and only 2 of 8 features were correlated (11111111 11222222 22333333 22444444). Here instance level structure among the *unsystematic* 6 features proved more predictive than the target rule; knowing the value of feature 3 allows perfectly reliable prediction of 5 other features, even though any such rule only applies to only one instance. The target features, overall, are less predictive and predictive in fewer rules. Thus, focused sampling is useful only when there is signficant structure available above the instance level.

These results might be glossed by the claim that focused sampling provided greater benefit for more complex (more instances, more features) and more "natural" (multiple correlated cues in a context of many uncorrelated ones, regularities beyond the instance) learning problems.

Summary

This research presents evidence and arguments for how learners use correlated structure to guide complex learning without feedback. Three experiments confirmed the predicted rule by rule facilitation from a more structured context (clustered feature facilitation). A mechanism, focused sampling, was outlined for using correlational structure to guide rule selection by altering the importance of features. The results of simulating the focused sampling procedure 1) produced clustered feature facilitation as predicted informally and as demonstrated by subjects and 2) suggested that focused sampling may be a particularly beneficial strategy for complex natural learning without feedback.

References

Billman, D.O. (1983). *Procedures for learning syntactic categories: A model and test with artificial grammars*. Doctoral dissertation, University of Michigan. Ann Arbor, Michigan: University Microfilm.

Billman, D.O. & Heit, E. (1987). Observational Learning From Internal Feedback: A simulation of an adaptive learning method. in review.

Fisher, M.A., & Zeaman, D. (1973). An attention-retention theory of retardate discrimination learning. In N.R. Ellis (Ed.), *International Review of Research in Mental Retardation*. New York: Academic Press.

Green, T.R.G. (1979). The necessity of syntax markers: Two experiments with artificial languages. *Journal of Verbal Learning and Verbal Behavior*, *18*, 481-496.

Lovejoy, E. (1966). Analysis of the overlearning reversal effect. *Psychological Review*, *73*, 87-103.

Maratsos, M.P. & Chalkley, M.A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K.E.Nelson (Ed.), *Children's Language*. New York: Gardner Press Inc.

Medin, D.L. (1983). Structural principles in categorization. In T.T.Tighe & B.E.Shepp (Eds.), *Perception, Cognition, and Development*. Hillsdale, N.J.: Erlbaum Publishers.

Medin, D.L. & Schaffer, M.M. (1978). A context theory of classification learning. *Psychological Review*, *85*, 207-238.

Morgan, J.L. & Newport, E.L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior*, *20*, 67-85.

Rosch, E.H. (1978). Principles of categorization. In E.H. Rosch & B.B.Lloyd (Eds.), *Cognition and Categorization*. Hillsdale, N.J.: Erlbaum Publishers.

Trabasso, T., & Bower, G.H. (1968). *Attention in Learning*. New York: Wiley.

Zeaman, D. & House, B.J. (1963). The role of attention in retardate discrimination learning. In N.R. Ellis (Ed.), *Handbook of Mental Deficiency*. New York: McGraw-Hill.

Notes

[1]The research was supported by BRSG RR-07083-19, a University of Pennsylvania Research Foundation Grant, and NIMH grant R23HD20522-01A1. We would like to thank Anne D'Ulisse and Valerie Sessa for their help in running the experiments.

Address for correspondence: Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, PA 19104.