

Isolated and interrelated concepts

ROBERT L. GOLDSTONE

Indiana University, Bloomington, Indiana

A continuum between purely isolated and purely interrelated concepts is described. Along this continuum, a concept is interrelated to the extent that it is influenced by other concepts. Methods for manipulating and identifying a concept's degree of interrelatedness are introduced. Relatively isolated concepts can be empirically identified by a relatively large use of nondiagnostic features, and by better categorization performance for a concept's prototype than for a caricature of the concept. Relatively interrelated concepts can be identified by minimal use of nondiagnostic features, and by better categorization performance for a caricature than for a prototype. A concept is likely to be relatively isolated when subjects are instructed to create images for their concepts rather than find discriminating features, when concepts are given unrelated labels, and when the categories that are displayed alternate rarely between trials. The entire set of manipulations and measurements supports a graded distinction between isolated and interrelated concepts. The distinction is applied to current models of category learning, and a connectionist framework for interpreting the empirical results is presented.

Modern research on concept representation and learning has evolved from two traditions. One tradition connects concept acquisition with language in general and word learning in particular (Lakoff, 1987; Saussure, 1915/1959). Concepts are approximately equated with single words or phrases. In this tradition, for example, evidence that a child has acquired the adult concept of *dog* comes from the child's use of the word "dog" to designate dogs. The other tradition connects concept acquisition with object recognition (Biederman, 1987). From this perspective, concept learning involves learning to correctly categorize perceptual inputs into classes.

These two approaches—linking concepts to language or to perception—are not mutually exclusive, and a full account of the nature of conceptual representation will most likely borrow from both approaches. Still, these approaches emphasize different methods for representing and processing concepts. There are three aims in the present paper: to describe these different characterizations, to develop empirical methods for distinguishing between them, and to devise a framework that integrates both characterizations into a single formal framework.

Concepts appear to be isolated from each other, each acting as an independent detector polling the world, yet

they also seem to influence each other within an interacting network (Collins & Quillian, 1969). For example, while certain shapes seem to be directly recognized as examples of dogs, the concept *dog* also appears to be closely associated with and influenced by other concepts such as *mammal*, *tail*, and *domestication*. Most models of concept representation and use try to account for one or the other of these types of conceptual behavior. The central purpose of this paper is to make a case for considering a continuum of interrelatedness in concept characterizations. The characterization of a concept refers to the way in which the concept is described and to the members that belong to it. A concept's characterization depends both on its representation and on the cognitive processes that operate on that representation. A concept is interrelated with respect to another concept to the extent that its characterization is influenced by the other concept.

In the present paper, the proposed continuum of interrelatedness is supported by results of five experiments in which the degree of interrelatedness of concepts acquired in the laboratory by human subjects was systematically manipulated. Three manipulations of conceptual interrelatedness and two empirical indicators of conceptual interrelatedness will be evaluated. In addition, a computational simulation of concept learning will be presented. The simulation demonstrates how degrees of conceptual interrelation can be modeled in a formal network. In the model, the use of a concept can reflect intermediate degrees of interrelation, depending on the strength of the connections that other concepts have to it. Such links are potentially determined by both the context of concept acquisition and the nature of the current processing demands.

Interrelated Concepts

There are several ways in which the characterization of concepts might be influenced by other concepts (Goldstone, 1991). First, a concept might be characterized as a

Experiments 1 and 2 were presented at the Thirteenth Annual Conference of the Cognitive Science Society, University of Chicago, August 1991. Experiment 4 was presented at the Fifteenth Annual Conference of the Cognitive Science Society, University of Colorado, June 1993. The author wishes to express thanks to Dorrit Billman, Michael Gasser, John Kruschke, James Hampton, Ben Martin, Douglas Medin, Robert Nosofsky, Steven Ryner, Richard Shiffrin, Edward Smith, and Linda Smith for their criticisms and suggestions, and to Paula Niedenthal for extensive help with editing and ideas. The research was supported by National Science Foundation Grant SBR-9409232. Requests for reprints should be sent to R. Goldstone, Psychology Department, Indiana University, Bloomington, IN 47405 (e-mail: rgoldsto@indiana.edu; further information can be found at <http://cognitrn.PSYCH.indiana.edu/>).

modified version of another previously acquired concept. Such modifications could include adding or deleting properties, increasing or decreasing the amount of a property, or rearranging properties. For example, one's concept of *toy poodle* might be a modified version of one's concept of *standard poodle*, with a modification to size. Or consider the concept *unicorn*. At first pass, one might think of unicorns as horses, with the addition of the features *magical* and *horns*. If *unicorn* is characterized by its relation to the concept *horse*, then altering one's concept of horse would likely also alter one's concept of *unicorn*. For example, if a new fact were learned about horses (e.g., the pad in the middle of a horse's hoof is called a "frog"), it might be incorporated into the *unicorn* concept as well (Potts, St. John, & Kirson, 1989).

A second way in which concepts can be influenced by other concepts is suggested by the work of Barr and Caplan (1987; Caplan & Barr, 1991). These researchers report evidence from feature-listing tasks that people's concepts typically contain both intrinsic and extrinsic features. Intrinsic features refer to parts or properties of the concept under scrutiny. Extrinsic features are "represented as the relationship between two or more entities" (p. 398). For example, an extrinsic feature of *hammer* is that it is "used to strike nails." This feature refers to a concept other than *hammer*. If this feature is part of one's concept of *hammer*, one cannot possess a *hammer* concept without also possessing *nail*.

In the preceding examples of interrelatedness among concepts, the *intension*, or internally represented meaning (see Johnson-Laird, 1983), of a concept is influenced by other concepts in the system. The *extension* of a concept is the set of items that is covered by the concept. In arguing for complete interrelatedness among concepts, Saussure (1915/1959) stated that "concepts are purely differential and defined not in terms of their positive content but negatively by their relations with other terms in the system" (p. 117). That is, he contended that all concepts are "negatively defined," or defined solely in terms of other concepts. Several current theories of word and concept representation from linguistics and computer science similarly assume that concepts' meanings are interrelated. In semantic network representations (Collins & Quillian, 1969), concepts are represented by their relations (e.g. *part-of*, *is-a*, and *has-a* relations) to other concepts. For example, the meaning of *cat* is represented by its *has-a* relation to *claw*, its *is-a* relation to *pet* and *animal*, and its *part-of* relation to *household* (see McNamara and Miller, 1989, for a review of evidence for theories of conceptual representation that posit interrelations between concepts).

In the experiments and simulations reported here, not all varieties of interrelated concepts were evaluated. Concepts interrelated by their semantic associations were not considered, nor were concepts interrelated by their cohabitation in a single theory or system (Fodor, 1983). Rather, the research was focused on artificial, laboratory-created concepts that were interrelated because they were

mutually exclusive categories; that is, membership in one category prevented membership in the other. Although visual concepts were used in these experiments to be reported, this relation of mutual exclusivity between concepts has been discussed by linguists. In linguistic theory, a mutually exclusive set of terms that are organized under an inclusive covering term is called a *contrast set* (Grandy, 1987; Lehrer & Kittay, 1992; Martin & Billman, 1994). For example, *Sunday*, *Monday*, and *Tuesday* belong to a common contrast set: days of the week. Concepts within a contrast set evoke each other, and once evoked, they influence each others' interpretation. In Saussure's example (1915/1959), *mutton* is influenced by the presence of other neighboring concepts. *Mutton's* use does not extend to refer to sheep that are living, because there is another concept that covers living sheep (*sheep*), and *mutton* does not extend to refer to cooked pig because of the presence of *pork*.¹ If one did not possess the concept *mutton*, "all its content would go to its competitors" (Saussure, 1915/1959, p. 116) and cooked sheep would be encompassed by *sheep*. Thus, concepts compete to the extent that they are conceptually related "neighbors." In empirical support of this notion, Brownell and Caramazza (1978) found that the ease of applying a term (e.g., *high*) to a stimulus depended on whether competing terms (e.g., *very high*) were also available to a subject.

Another source of linguistic evidence for competition between concepts comes from work on the mutual exclusivity hypothesis (Markman, 1990; Waxman, Chambers, Yntema, & Gelman, 1989). According to this work, children determine the referent of a noun by assuming that nouns are mutually exclusive, and consequently, if a new term is applied to one of two objects, and if one of these objects already has a name, children will tend to assume that the term refers to the other object. In this example, as with the Saussurian notion of competition between concepts for control of conceptual regions, the influence of concepts upon each other is negative. As one concept gains control of a conceptual area, its competitor concepts lose control of the area. However, concepts can also be positively interrelated if they are inductively or hierarchically related to each other (Waxman & Senghuas, 1992). Properties that are found to be true of one concept will frequently be transferred to similar concepts (Osherson, Smith, Wilkie, Lopez, & Shafir, 1990), more general concepts, and more specific concepts.

As with neighbor-influenced concepts, several researchers have suggested that categories are constructed so as to be highly differentiated from each other. Rosch and her colleagues (Rosch, 1975; Rosch & Mervis, 1981) contend that natural categories are created so that there is as little overlap as possible between members of different categories (and as much overlap as possible between members of the same category). Corter and Gluck (1992) and Anderson (1991) present formal accounts in which categories tend to be formed such that the prediction of features from knowledge of category membership is maximized. This constraint underlies the development of con-

cepts that contain highly distinct sets of instances (see also Krueger & Rothbart, 1990, and Rumelhart & Zipser, 1985, for different formalizations of competition between categories).

Isolated Concepts

Theoretical grounds for doubting the ubiquity of completely interrelated concepts derives from a reappraisal of Saussure's notion of competition between concepts. In his view, concepts compete with neighboring concepts to cover specific instances. In this way, concepts are completely defined by their neighbors. But how are the neighbors of a concept determined in the first place? It seems that concepts must usually have an independent characterization in order to have neighbors. Harnad (1990) and Barsalou (1993) have argued that concepts cannot simply gain their meaning from other concepts; concepts must also be grounded by perceptual, nonsymbolic properties (see also Goldstone, 1994b). In most cases, a concept can have neighbors only if it first has some location. This location is its isolated or concept-independent characterization. Such reasoning suggests that a good diagnostic for locating objects on the isolated/interrelated continuum is, "Would this concept be used in this way if some/most/all other concepts were eliminated?"

One way to conceive of an isolated concept is as a feature detector. In the classical conception of feature detectors, a detector becomes activated when an input with a particular feature is displayed. For example, there exist neurons in primate striate cortex that act as feature detectors for straight lines of particular orientations (Hubel & Wiesel, 1968). There is also evidence that some neurons respond selectively to stimuli as complex as triangles, hands, and faces (Bruce, Desimone, & Gross, 1981). In order for a neuron to become active when a line of a particular orientation is presented to the visual field, the feature detector does not need any information from other detectors, concepts, or theories in the system. Although the particular specialization that a feature detector develops can be influenced by the states of other neighboring neurons at the same cortical level (Malsburg, 1973), once the specialization of a neuron has been set, the feature detector is isolated from the influences of other neurons for the most part.

The lately unfashionable (but see Ullman, 1989) template approach to pattern recognition provides another possible representation for isolated concepts. If patterns are categorized by their being compared with stored category templates, the representation of the category does not depend on other category representations. A category's representation is simply the image that is compared with the input to be recognized. For example, instead of characterizing *unicorn* by making reference to *horse*, one's representation may simply be a photograph-like image.

Researchers interested in concept learning from the perspective of categorization of perceptual stimuli have proposed prototype and exemplar representations. Posner and Keele (1968) and Reed (1972) advocate prototype representations. Medin and Schaffer (1978) and No-

sofsky (1986) argue that concepts are represented by their individual exemplars. At a first pass (see the General Discussion for a second pass), these internal representations do not depend on other concepts. That is, the prototype or exemplars that create a category's representation are not typically altered by other categories.

EMPIRICAL METHODS FOR ANALYZING CONCEPTUAL INTERRELATEDNESS

The majority of concepts are probably neither purely isolated nor purely interrelated. The research of Harnad, Barsalou, and psychologists investigating concept learning points to the insufficiency of theories that define concepts solely in terms of other concepts. However, there is also substantial evidence for rich interconceptual relations (e.g., Winston, Chaffin, & Herrmann, 1987). The empirical methods described here were aimed at locating concepts along the isolated-to-interrelated continuum. They diagnose only the subset of interrelated characterizations wherein a concept is influenced by competition from other simultaneously acquired concepts.

Overall, the present research strategy was to use converging operations (Garner, Hake, & Eriksen, 1956) to identify isolated and interrelated concepts. Experimental manipulations were developed that were expected to alter the interrelatedness of the concepts to be learned. At the same time, empirical indicators of interrelatedness were also developed. The reason for testing multiple manipulations and indicators is that each indicator, by itself, provides only an indirect and fallible lens on interrelatedness. Confidence that a particular manipulation or indicator is related to interrelatedness increases if it provides results that are consistent with the other methods. Considered individually, what each manipulation is affecting and what each indicator is measuring is not precisely specifiable. When the manipulations and indicators are considered together, the pattern of results strongly suggests a common underlying construct of concept interrelatedness. Table 1 presents an overview of the converging manipulations and indicators that were explored.

The present experiments focused on three methods for experimentally manipulating concept interrelatedness, and on two methods for measuring interrelatedness. The notion that concepts vary systematically in their interrelatedness would be supported if the experimental manipulations and measures consistently cohere together in locating particular concepts on the continuum of interrelatedness. Interrelatedness was manipulated by (1) instructing subjects to either create independent images of the concepts to be learned (isolated) or look for features that distinguish the concepts from each other (interrelated), (2) alternating presented concepts frequently (interrelated) or rarely (isolated), and (3) giving the concepts related (interrelated) or unrelated (isolated) labels. Concept interrelatedness is measured by the degree to which nondiagnostic features are used for categorizing and depicting categories, and by the relative ease of categorizing instances that are prototypical or distorted in

Table 1
Overview of Manipulations and Diagnostics

Manipulation	Prediction	Diagnostic	Indicators
Instructions (Experiments 1, 4)	imagery = isolated discriminate = interrelated	Influence of nondiagnostic features on accuracy (Experiments 1, 2, 3)	large influence = isolated little influence = interrelated
Category alternation (Experiments 2, 5)	seldom = isolated often = interrelated	Presence of nondiagnostic features in drawings (Experiments 1, 2, 3)	frequent appearance = isolated rare appearance = interrelated
Category labels (Experiment 3)	standard = isolated negation = interrelated	Categorization speed of caricature and prototype (Experiments 4, 5)	prototype advantage = isolated caricature advantage = interrelated
Practice (Experiments 2, 3, 4, 5)	early = isolated late = interrelated	Degree of caricaturization in subject-made drawings (Experiments 4, 5)	low degree = isolated high degree = interrelated
Degree of distortion (Experiments 1, 2, 4, 5)	low = isolated high = interrelated	Influence of within and between category similarity	large w/i influence = isolated large b/w influence = interrelated
Category frequency	frequent = isolated rare = interrelated	Dissemination of information to new concepts	restricted = isolated disseminated = interrelated
Presentation order	first category = isolated second category = interrelated		
Classification task	Go/no-go = isolated classification = interrelated		

particular ways. The justification for associating each of these manipulations and measures with concept interrelatedness will be discussed as they are introduced.

EXPERIMENT 1

Nondiagnostic Features in Categorization

One method for identifying isolated and interrelated concepts is to observe the influence of nondiagnostic features on categorization accuracy. A nondiagnostic feature is a feature that does not, by itself, provide any information to choose between candidate categories. When one categorizes a shape as a triangle or a square, knowledge that the shape has the feature “straight line” does not provide any grounds for choosing between the categories. If a feature F is nondiagnostic, then $P(C|F) = P(C)$ for all categories C ; that is, the probability of a category, given that the feature is present, is equal to the probability of the category, given no additional information. The category validity of a feature is $P(F|C)$, the probability of the feature, given a particular category. Although “straight line” is not diagnostic for choosing between a triangle and square, the feature is certain [$P(F|C) = 1.0$] for both categories.

The difference between a diagnostic and a nondiagnostic feature is only relevant for interrelated concepts. For purely isolated concepts, any feature that has high category validity will likely be a part of the template, or image, of the concept and may be used for categorization purposes. Furthermore, if the categorization decision rule involves a similarity criterion, then nondiagnostic features can increase categorization accuracy. An example of such a rule is: “If the similarity of the item to the prototype of Category Y is greater than level X , then place the item into Category Y . If the item’s similarity to any category never exceeds X , then respond at random.”

Such rules are commonly used in pattern recognition domains. Nondiagnostic features can increase categorization accuracy, because they cause an item’s similarity to a category to exceed the criterion X .

The diagnosticity of a feature only becomes an issue when there is a set of candidate categories and an attempt is being made to *distinguish* between the possible choices. Nondiagnostic features, by definition, do not distinguish between candidate categories and would not be included in a concept that is characterized purely by its negative relation to its competitor concepts. The experimental question then becomes: How much does the presence of nondiagnostic features with high category validity increase the accuracy of categorization? For relatively isolated concepts, nondiagnostic features should increase accuracy because they will be part of the concepts’ characterization—assuming that the probability of finding the correct category for an item increases as the item’s similarity to the correct category’s template increases (see General Discussion for details regarding this assumption). For relatively interrelated concepts, nondiagnostic features should have a small influence on accuracy because they will not be a large part of the concept’s characterization. For concepts that are influenced by competition with another concept, features that discriminate between concepts will play a large role.

Overview of Experiment

In Experiment 1, the influence of nondiagnostic relative to diagnostic features was used as the indicator of interrelatedness. An instructional manipulation was used in an attempt to vary concept interrelatedness. One group of subjects was told to create an image of the two concepts to be learned, and another group of subjects was told to seek out stimulus features that served to distinguish the concepts. The *image* instructions were aimed at promot-

ing isolated concepts; if an image/template is formed for each concept, there should be relatively little influence of one concept on another concept's characterization. The second set of instructions was aimed at promoting interrelated concepts. A concept's distinguishing features are only diagnostic *relative to another concept*. While an image can be generated for a concept without any knowledge of the other concepts being acquired, the selection of distinguishing/diagnostic features for a concept requires knowledge of the other candidate concepts.

Thus, the prediction for the first experiment was that subjects who were given "image" instructions would develop concepts that were relatively isolated, as indicated by a large influence of nondiagnostic features on categorization accuracy. In addition to the categorization accuracy measure, clues about the nature of a concept's representation were also obtained by asking subjects to pictorially describe their concepts.

Method

Materials. Sample materials are shown in Figure 1. The stimuli were composed of the 20 horizontal, vertical, and diagonal line segments that connected a 3×3 grid of dots. The line segments, but not the dots, were displayed to subjects. The background for all the materials was white. For each of the two concepts to be learned, a prototypical pattern of line segments was generated, consisting of 9 black lines. Out of the 20 positions for line segments, 10 of the positions were diagnostic and 10 were nondiagnostic. A line segment was diagnostic if it provided evidence in favor of one of the concepts—that is, if it was present in one concept prototype but not the other. A line segment was nondiagnostic if its presence did not make one concept more likely than the other. White (absent) line segments could also be nondiagnostic or diagnostic, although postexperimental interviews revealed that they were not as salient as the black line segments for most subjects. Most analyses were unchanged when they were excluded as features. Out of the 10 diag-

nostic line segments, 5 were black and 5 were white; out of the 10 nondiagnostic line segments, 4 were black and 6 were white.

Subjects were presented with distortions of the two prototypes. Three possible distortions of Category A are shown in Figure 1. Distortions were formed by randomly switching (from white to black, or from black to white), with a probability of .2, each of the 20 line segments of the concept prototypes. Thus, on the average, distortions contained 80% of the line segments of their prototypical pattern. Individual nondiagnostic and diagnostic line segments were switched equally often, and the two prototypes were used as the basis for the distortions equally often. Thus, the category validity [$P(\text{cue} | \text{category})$] of all line segments in the prototype was .8, the diagnosticity [$P(\text{category} | \text{cue})$] of diagnostic line segments was .8, and the diagnosticity of nondiagnostic line segments was .5.

Procedure. Twenty-eight undergraduates from the University of Michigan were divided into two groups. One group of subjects (the *image* group) was given the instructions, "While you are learning the two categories, you should try to form an *image* of what each category looks like. Use these images to help you categorize the pictures that you see." The other group of subjects (the *discriminate* group) was given the instructions, "While you are learning the two categories, you should try to find features in the pictures that help you *distinguish* between the two categories."

Six hundred randomly generated distortions of the two prototypes were displayed. On each trial, a distortion of a concept's prototype was displayed, and subjects pressed one of two keys to indicate their categorization decision. The image remained on the screen until the subject entered a response. Subjects then received feedback indicating whether their choice was correct and what the correct response was. The feedback was displayed for 1.5 sec, and a blank screen between trials was displayed for 1 sec. All materials were presented on Macintosh SE computers.

Subjects were given breaks after every 100 trials. During the break, subjects were reminded to either form images or look for discriminating features. Following the categorization task, subjects were instructed to "draw pictures that best capture the nature of the two concepts" on empty 3×3 grids.

Results

The data of principal interest concern the use of diagnostic and nondiagnostic features by subjects in the two instruction groups. The number of diagnostic and nondiagnostic features that were altered from a concept's prototype was measured on each trial. Figure 2 illustrates the unsurprising finding that categorization performance decreased as the number of altered line segments increased [$F(4,104) = 17.5$, $MS_e = 0.18$, $p < .01$]. Separate results are shown for alterations of nondiagnostic and diagnostic features, for both instruction groups. For example, the 85% accuracy rate for discriminate instructions when no diagnostic features were changed includes all of the stipulated data, regardless of the number of nondiagnostic features altered.

In addition to the main effect of the degree of distortion on accuracy, altering diagnostic features was more detrimental to categorization accuracy than was altering nondiagnostic features. The primary method of measuring the influence of a particular type of features was to perform a regression of the number of features altered on categorization accuracy for each subject. The slope of the resulting regression equation was positively related to the influence of the features. The slopes of the lines relating number of altered features to categorization accuracy, -5.2 for diagnostic lines and -2.3 for nondiag-

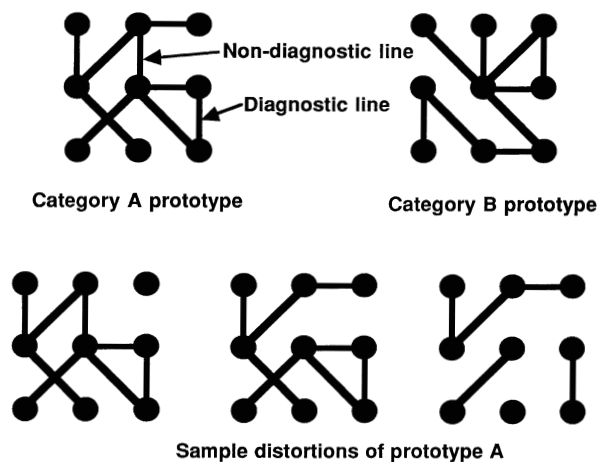


Figure 1. Sample stimuli from Experiment 1. Nondiagnostic line segments were included in both categories' prototypes. Diagnostic line segments were included in only one category's prototype. Distortions of categories were produced by randomly altering line segments with a probability of .2. In the actual materials shown to subjects, only the line segments and not the dots were displayed.

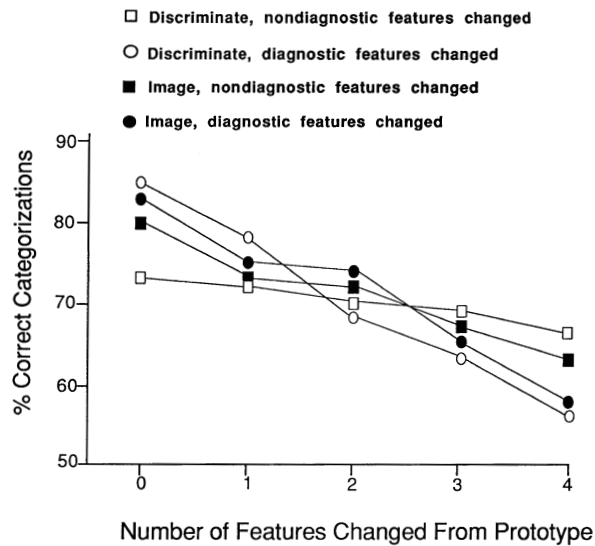


Figure 2. Results from Experiment 1, showing an interaction between instructions and type of features changed on categorization accuracy. The influence of a type of feature is given by the slope of its line.

nostic lines, were significantly different [$F(1,26) = 8.8$, $MS_e = .09$, $p < .01$].

The result of principal interest was the significant three-way interaction between segment diagnosticity, number of features changed, and instruction type [$F(4,104) = 3.59$, $MS_e = 0.21$, $p < .05$]. When subjects were given image instructions, nondiagnostic features were more important than when subjects were given discriminate instructions. When subjects were given discriminate instructions, categorization accuracy was only slightly affected by alterations to nondiagnostic features. These nondiagnostic features had a greater influence when subjects were told to form images of the concepts.

The influence of a type of feature was also quantified by measuring the slope of the line that related the number of alterations of a particular type of feature to categorization accuracy. For the discriminate group, slopes were -1.4 and -5.6 for nondiagnostic and diagnostic features, respectively. For the image group, the respective slopes were -3.2 and -4.8 . These slopes reveal a significant interaction between instruction group and feature diagnosticity on accuracy [$F(1,26) = 5.45$, $MS_e = 0.30$, $p < .05$]. For both groups, even nondiagnostic features had a significant influence on categorization accuracy, as measured by comparing the obtained slopes to a population mean of 0.0 [$F(1,26) > 6.15$, $MS_e < 0.16$, $p < .05$].

The relative influence of nondiagnostic and diagnostic features did not remain constant across training. There was a marginally significant trend in which concepts in the image group became increasingly isolated with practice, as indicated by a marginally significant interaction between trial number and feature diagnosticity on the slope of the line relating number of altered features to categorization accuracy [$F(1,13) = 4.4$, $MS_e = .21$, $p < .06$]. For the first half of the trials, the influence of non-

diagnostic features (again quantified by the slopes of the lines relating number of feature changes to categorization accuracy) was 49% of the influence of diagnostic features (nondiagnostic slope = -1.73 , diagnostic slope = -3.53). For the second half of trials, the influence of nondiagnostic features rose to 77% of the influence of diagnostic features (nondiagnostic slope = -4.67 , diagnostic slope = -6.07). In contrast, concepts in the discriminate group became more interrelated with practice [$F(1,13) = 5.50$, $MS_e = 0.25$, $p < .05$]. For the first half of the trials, the influence of nondiagnostic features was 33% of the influence of diagnostic features (nondiagnostic slope = -3.57 , diagnostic slope = -10.81). For the second half of trials, nondiagnostic features were only 10% as influential as diagnostic features (nondiagnostic slope = -0.039 , diagnostic slope = -0.386).

An examination of the pictures drawn by subjects to exemplify concepts provided additional evidence for different characterizations developing in the two instruction groups. Subjects' pictures were analyzed in terms of the numbers of diagnostic and nondiagnostic features correctly depicted (i.e., part of the concept's prototype). Image instructions yielded a marginally significant higher proportion of correctly drawn features (an average of 14.6 out of 20 correct segments) than did discriminate instructions (13.9 correct segments) [$t(26) = 1.94$, $p = .06$]. Importantly, the difference was particularly large for nondiagnostic features (image = 8.2 out of 11 correct segments, discriminate = 6.7 correct) [$t(26) = 2.25$, $p < .05$].

Discussion

Experiment 1's results were predicted by the following set of assumptions: (1) nondiagnostic features are more likely to influence categorization accuracy for relatively isolated than for relatively interrelated concepts, (2) imagery, as opposed to discrimination, instructions are more likely to promote isolated concepts, and (3) the individual line segments are the basic features of the stimuli. Discussion and justification of the first two assumptions will be postponed until the General Discussion. A possible objection to the third assumption is that individual line segments may not be the correct unit of analysis. There is no a priori reason why the psychologically relevant features must coincide with the experimenter-defined features. Subjects might develop features that are compositions of several line segments (Hock, Tromley, & Polmann, 1988; Palmer, 1978; Schyns, Goldstone, & Thibaut, in press). For example, subjects may encode Concept A in Figure 1 as possessing the feature "X-shape in the lower left quadrant." This feature is diagnostic in that B examples will not typically have this feature. Furthermore, by taking away a "nondiagnostic" line, we eliminate this diagnostic feature. According to this objection, nondiagnostic line segments may influence categorization accuracy only because they serve to define psychologically salient, diagnostic, complex features.

However, the obtained results provide a provisional validation of line segments as important functional fea-

tures in some cases. Under some circumstances, subjects' analyses of features clearly did coincide with the experimenter-determined analysis. Specifically, when subjects were instructed to look for discriminating/diagnostic features, the lines that were experimentally defined as nondiagnostic did not have much influence on categorization accuracy relative to diagnostic lines. The nondiagnostic line segments never lost their influence completely, and this can be taken as some support for an influence of complex features that are constructed from simpler units. Simply arguing that subjects' features may not have agreed with the experimenter's features does not explain why sometimes the two did agree. The isolated/interrelated conceptual analysis that has been developed here thus gains some support because it provides an account for when and why the experimenter's and subject's features coincide at the level of individual line segments. Although an alternative "complex features" account for the effect of the experimental manipulation in Experiment 1 is possible, later experiments will cast doubt on the generality of this explanation.

The data from the subjects' drawings of the concepts are potentially important, because they provide relatively direct information about the concepts' characterizations. The forced-choice nature of categorization decisions creates a heavy bias for relational judgments; categorizations will often be based on the relative appropriateness of one category over another. Subjects' drawings, on the other hand, are absolute measures of the qualities of one category irrespective of the other. Subjects from the image group, posited to produce concepts that were relatively isolated, were more likely to include nondiagnostic features in their visual depictions of concepts than were subjects who were told to find distinguishing features. These results provide converging evidence for the categorization accuracy results.

EXPERIMENT 2

Experiment 2 was an attempt to obtain evidence converging with Experiment 1, using a second task manipulation intended to vary the interrelatedness of concept representations. In this experiment, the frequency of category alternation was varied. The presentation of different categories was alternated frequently or rarely. If categories are alternated rarely, subjects will see long clusters of items that belong to the same category; a subject may, for instance, see six pictures that belong to Category 1 followed by five pictures belonging to Category 2. If categories are alternated frequently, then most often a Category 1 picture will be followed immediately by a Category 2 picture, and vice versa.

The a priori assumption for this experiment is that frequent alternation of categories will yield concepts that are interrelated, and infrequent alternation will yield isolated concepts. The justification for this assumption is that if several distortions of the same concept's prototype are presented in a series, the concept's characterization

will most likely be based on the common properties of the distortions. Influences from the other concept will be relatively modest. If categories are alternated frequently, there will be more opportunity for interplay between the developing concepts. In short, an item's categorization is likely to depend disproportionately on its immediate context; if that context contains members from other categories, the item will likely be encoded in terms of those other category members (Medin & Edelson, 1988). If its context contains members from the same category, the item will likely be encoded in terms of its overlap with these items. Some empirical justification for this hypothesis comes from work on children's word learning suggesting that words that are presented in the immediate context of each other tend to be given mutually exclusive definitions more than when they are presented sequentially (Waxman et al., 1989).

Method

Materials. The same type of materials used in Experiment 1 was used here. Two concept prototypes were created that each had 8 diagnostic and 12 nondiagnostic line segment positions.

Procedure. Twenty-six Indiana University undergraduates were given the same categorization task used in Experiment 1. Subjects were given no strategy instructions; they were simply instructed to select a category for each of the presented pictures. The only other departure from Experiment 1's procedure was that categories were alternated frequently for half of the subjects, and rarely for the other half. For both groups, Category 1 and Category 2 items were each presented on one half of the trials. For the frequent alternation group, the probability of displaying an item from the same category as that of the preceding item was .25. For the infrequent alternation group, the probability was .75.

Results

The results from Experiment 2 showed a similar pattern to those from Experiment 1, as presented in Figure 3. There is a significant three-way interaction between diagnosticity, number of features altered, and alternation frequency [$F(4,96) = 3.2$, $MS_e = 0.31$, $p < .05$]. Diagnostic line segments influenced categorization accuracy more than did nondiagnostic lines, as measured by the slopes of lines relating the number of features altered from the prototype to categorization accuracy (nondiagnostic slope = -1.5 , diagnostic slope = -2.8) [$F(1,24) = 14.7$, $MS_e = 0.15$, $p < .05$]. More importantly, the differential influence of diagnostic lines was greater when the category of the displayed item was alternated frequently rather than infrequently, as measured by the slopes (nondiagnostic frequent slope = -1.2 , nondiagnostic infrequent slope = -1.8 , diagnostic frequent slope = -4.0 , diagnostic infrequent slope = -2.8) [$F(1,24) = 5.79$, $MS_e = 0.12$, $p < .05$].

A similar pattern of results was obtained from subjects' drawings. Subjects who received frequently alternating concepts drew an average of 6.5 correct diagnostic lines (out of 9 possible), and 7.3 correct nondiagnostic lines (out of 11 possible). Subjects who received infrequently alternating concepts drew an average of 6.7 correct diagnostic lines and 8.5 correct nondiagnostic lines. These

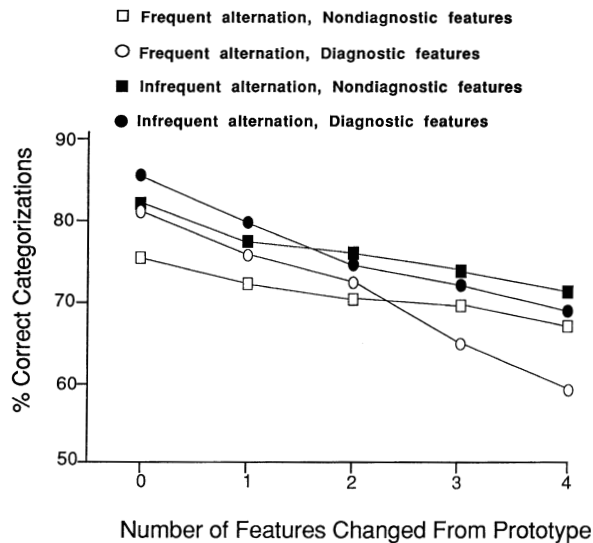


Figure 3. Results from Experiment 2, showing an interaction between frequency of category alternation and type of features changed on categorization accuracy. The influence of a type of feature is given by the slope of its line.

results indicate that significantly more nondiagnostic features were drawn as part of a concept's representation when categories were infrequently alternated [$F(1,24) = 5.58$, $MS_e = 1.2$, $p < .05$].

Categorization accuracy was significantly higher for infrequently altered categories than for frequently altered categories [$F(1,24) = 8.3$, $MS_e = .09$, $p < .05$]. With increasing practice, both groups of subjects' concepts became increasingly interrelated, as measured by a significant blocks (first half, second half) \times feature (nondiagnostic, diagnostic) interaction with slopes as the dependent measure [$F(1,24) = 4.4$, $MS_e = 0.10$, $p < .05$]. The slopes representing the influence of diagnostic features were -3.2 and -3.8 early and late in training, respectively [$F(1,24) = 1.2$, $MS_e = 0.25$, $p > .1$]. The slopes for early and late nondiagnostic features were -2.0 and -0.8 respectively [$F(1,24) = 5.0$, $MS_e = 0.11$, $p < .05$].

Discussion

Experiment 2 provides support for a coherent set of manipulations and measures of concept interrelatedness. Similar categorization accuracy and concept depiction results were obtained when subjects were asked to depict images of concepts (Experiment 1) and when displayed categories alternated infrequently. The similar pattern of results was predicted by the hypothesis that both of these manipulations encourage the creation of isolated concepts—concepts that are not influenced by the other concepts that are being acquired. Similarly, frequent category alternation produced results that were similar to those elicited by instructions to attend to distinctive features. Both of these manipulations were hypothesized to yield interrelated concepts—concepts with representations that are influenced by the other learned concepts.

The finding that categorization accuracy is greater for infrequently alternated categories than for frequently alternated categories corroborates results from Whitman and Garner (1963). Each of the two alternation conditions has an advantage over the other. Within the isolated/interrelated framework, frequent alternation of categories has the advantage of highlighting features that serve to distinguish categories. Conversely, infrequent alternation of categories has the advantage of highlighting information that remains constant across the members within a category (Hunt & McDaniel, 1993; Medin, Wattenmaker, & Michalski, 1987). Infrequent alternation may have resulted in better categorization performance because there were no features that perfectly distinguished the categories, several features were completely nondiagnostic, and the materials allowed for the relatively efficient creation of image-like category characterizations. Subjects who viewed several examples of a category successively may have adopted the efficient strategy of discovering frequent commonalities between the examples rather than trying to discover partially diagnostic features that distinguished the categories from each other.

EXPERIMENT 3

Until now, we have considered manipulations that influence the isolation/interrelation of both concepts to be acquired. It is also possible to devise asymmetric manipulations—manipulations that produce concepts that are not equally interrelated. One of the most straightforward ways to do this is to assign different labels to the concepts to be learned. In the *asymmetric* condition of Experiment 3, concepts were essentially labeled “Concept A” and “Not Concept A,” whereas in the *symmetric* condition, the two concepts were labeled “Concept A” and “Concept B.” Within the asymmetric condition, the concept labeled “Not Concept A” is predicted to be more influenced by “Concept A” than the concept labeled “Concept A” is influenced by “Not Concept A” (Clark, 1990). The concept that has a label that refers to another concept is predicted to be highly influenced by the referenced concept. Although logically identical categories were used in the two asymmetric groups and in the symmetric condition, the labels themselves were predicted to influence the degree of interaction between the acquired concepts. In addition, the “Not Concept A” concept is predicted to be more interrelated to its competitor concept than either concept in the symmetric labels condition, because concepts in the symmetric labels condition receive labels that are independent of each other. Some grounds for predicting an influence of labeling comes from work on hypothesis testing (Van Wallendaal & Hastie, 1990). Information which is presented as “A is not guilty” influences judgments of “A is guilty” far more than it influences “B is guilty” judgments, even when A and B are the only suspects. That is, when the negative contingency between two hypotheses is explicit in their labeling, the two hypotheses influence each other more

than when the two hypotheses are simply mutually exclusive (Goldstone, 1993a, discusses a related labeling bias). More generally, the efficacy of labeling manipulations in inducing changes to learned concepts has been shown by Wisniewski and Medin (1994).

Method

The same materials and procedures as those used in the previous experiments were used in this experiment, with a few exceptions. Twenty-four Indiana University undergraduates were split evenly into two groups: symmetric and asymmetric labeling. Subjects in the symmetric-labels group were instructed, "You will see abstract paintings from two different imaginary artists, Yarpleaux and Noogan. If you think a painting was created by Yarpleaux, press the 'Y' key, and press the 'N' key if you think it was painted by Noogan." Subjects in the asymmetric-labels group were instructed, "You will see paintings by the imaginary artist Yarpleaux, and several forgeries by other artists. If you think a painting was created by Yarpleaux, press the 'Y' key. If you do not think the painting is authentic, press the 'N' key." As before, subjects received feedback about the correctness of their responses. Subjects in the symmetric-labels group were corrected with the phrase "No. This painting is a Yarpleaux [Noogan]." Subjects in the asymmetric-labels group were corrected by the phrase "No. This painting is NOT a Yarpleaux" or by "No. This painting IS a Yarpleaux."

The particular prototypes that were labeled Yarpleaux, Not Yarpleaux, and Noogan were counterbalanced. Thus, for half of the subjects in the asymmetric-labels group, one prototype was labeled "Yarpleaux" and the other prototype was labeled "Not Yarpleaux." These labels were swapped for the other subjects.

Results

The results of principal interest, shown in Figure 4, pertained to the influence of nondiagnostic and diagnostic features on categorization accuracy within the different labeling conditions. The figure shows categorization accuracies as a function of the number of nondiagnostic or diagnostic features that were altered from the prototype. The control for the "Yarpleaux" labeling condition was the "Yarpleaux" condition from the symmetric group, and

the control for the "Not Yarpleaux" labeling condition was the "Noogan" condition. The difference in influence of nondiagnostic features between the two asymmetric-labels groups was greater than the difference between the symmetric groups, on the basis of the slopes of lines relating number of features altered (0, 1, 2, 3, or 4) and categorization accuracy [$F(1,22) = 5.5$, $MS_e = 0.22$, $p < .05$]. For the "Yarpleaux" category, the slope reflecting the influence of nondiagnostic features was -2.6 ; for the "Not Yarpleaux" category, the slope was -0.8 . For the respective control symmetric conditions, the slopes for nondiagnostic features were -1.3 and -1.7 . The same effect was not found for diagnostic features, as all four groups had approximately equal slopes ("Yarpleaux" = -6.2 , "Not Yarpleaux" = -4.8 , symmetric "Yarpleaux" = -5.9 , symmetric "Noogan" = -6.5) [$F(1,22) = 1.7$, $MS_e = 0.34$, $p > .1$]. The two symmetric groups never significantly differed for any analysis, and thus are combined together in Figure 4.

Marginally significant corroborating results were found when subjects were asked to draw their best representation of each of the concepts. Subjects drew an average of 6.3, 6.6, and 6.2 correct diagnostic lines [$F(2,22) = 1.1$, $MS_e = 0.87$, $p > .1$] and 7.9, 8.2, and 7.0 nondiagnostic lines [$F(2,22) = 3.1$, $MS_e = 1.1$, $p = .07$] for the symmetric, "Yarpleaux," and "Not Yarpleaux" conditions, respectively.

The influence of nondiagnostic features decreased [$F(1,22) = 5.9$, $MS_e = 0.35$, $p < .05$], and the influence of diagnostic features increased [$F(1,22) = 4.9$, $MS_e = 0.23$, $p < .05$], with practice, collapsing across labeling conditions. These results are consistent with those obtained in Experiment 2.

Discussion

Experiment 3 illustrates one method for creating asymmetrically interrelated concepts. When labeling con-

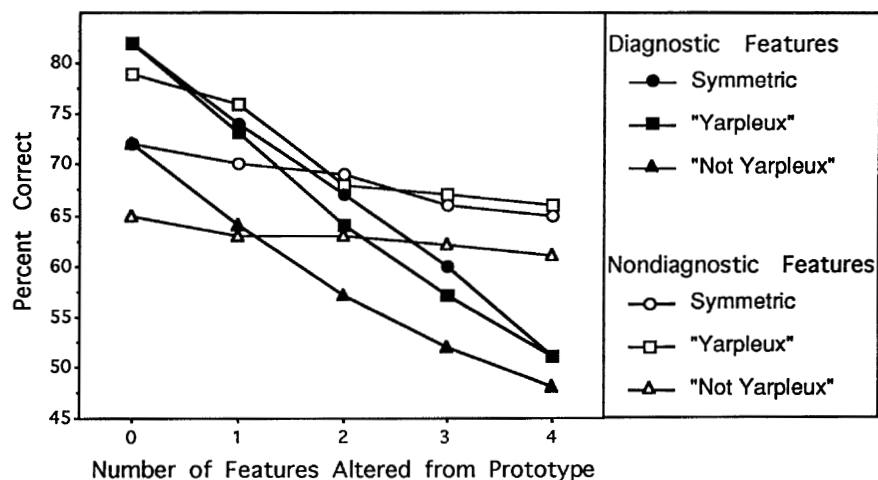


Figure 4. Results from Experiment 3, showing an interaction between the label given for a category and the type of features changed on categorization accuracy. The influence of nondiagnostic features on categorization accuracy is much greater for the positively labeled "Yarpleaux" category than for the negatively labeled "Not Yarpleaux" category.

cepts A and Not A, the A concept appeared to be relatively isolated and the Not A concept appeared to be influenced by the other concept, even though the actual prototypes were logically equivalent. Nondiagnostic features that were equally likely to occur for both A and Not A concepts were much more likely to be associated with A. When concepts were learned with labels that were not related to each other, they became represented in a fairly isolated manner.

The success of the labeling manipulation casts doubt on the generality of the alternative account for Experiment 1. The alternative account was that complex features that involve more than one line segment are more likely to develop with image than with discriminate instructions. Although this account is plausible for Experiment 1, for Experiment 3 there is little reason to think that complex features are more likely to develop for asymmetric positively labeled categories than for symmetric categories, or for symmetric categories than for negation-labeled categories.

Taken in total, the first three experiments show that concept interrelatedness, as indicated by reliance on nondiagnostic and diagnostic features, can be promoted by a number of experimental manipulations. Concepts are relatively interrelated when subjects are instructed to look for discriminating features, when categories are alternated often, and when categories are labeled as negations of other categories. Concepts are relatively isolated when subjects are instructed to create concept images, when categories are alternated infrequently, and when categories are given unrelated or standard labels.

PROTOTYPES, CARICATURES, AND CONCEPT DEFINITION

The second method for measuring the degree of concept interrelatedness involves comparing the categorization accuracy for a concept's prototype and caricature. Consider the simple case in which two concepts can be distinguished by their values on a single dimension. Concept A has instances that are 2, 3, and 4 in. wide, while Concept B has instances that are 6, 7, and 8 in. wide. The prototype of Concept A is the 3-in. item; the caricature of Concept A is the 2-in. item. The prototype of a concept is the item that possesses the central tendency of dimension values, averaging over all of the concept's instances. Caricatures of a concept assume dimension values that depart from the central tendency *in the opposite direction from the central tendency of other concepts to be simultaneously acquired*. Thus, 3 in. is the central tendency of Concept A because it is the average of 2, 3, and 4, but 2 in. is a caricature of Concept A because it is *less* than A's average of 3 in. and B's average of 7 in. is *greater* than A's average. Just as a caricature of a politician in a cartoon exaggerates certain distinguishing features of the politician, so a caricature of a concept makes the dimension value that distinguishes the concept from other concepts more extreme.

A question of interest is, Is the prototype or caricature of a concept more accurately categorized as belonging to the concept? On the one hand, it might be argued that the prototype will be better categorized. The prototype is the item that is closest to the most other category items, assuming either normal or uniform item distributions. According to prototype theorists, a concept's representation is based on the prototype of the concept, and the degree to which an item belongs to a concept is directly related to the item's proximity to the concept's prototype (Posner & Keele, 1968; Rosch, 1975). Some researchers have extended the notion of a prototype to include concepts that are defined in terms of an ideal or exaggerated item (Lakoff, 1987). However, in this discussion the notion of a prototype will be restricted to the central tendency or most average member of a category.

On the other hand, an argument can also be made that the caricature is better categorized. Several researchers have found a caricature advantage in categorization accuracy and/or speed (Rhodes, Brennan, & Carey, 1987; Nosofsky, 1991). Caricatures emphasize distinguishing category features (Rhodes et al., 1987) and are further removed from the boundary between categories than are prototypes (Ashby & Gott, 1988; Nosofsky, 1991).

The framework developed thus far predicts that whether a prototype or a caricature advantage is found will depend, in part, on the degree to which concepts are interrelated. If a concept is purely isolated, an advantage for the concept's prototype is expected. In the absence of interconceptual influences, the representation that best exemplifies a concept will be its prototype. However, if a concept is characterized relative to other concepts, rules of the sort "Concept A items are smaller than Concept B items" or "Concept A is small, relative to Concept B" will be likely to develop. Caricatures better fit these relational rules than do prototypes. While items of 2 or 3 in. both satisfy the rule "smaller than Concept B items," the item of 2 in. more clearly satisfies this rule. Thus, if a concept is purely isolated and categorization speed or accuracy is a function of the proximity of the item to the concept's best representation (Rosch & Mervis, 1975), the prototype is expected to be more accurately or quickly categorized than caricatures. Whether an item is a caricature or simply a distortion of a concept's prototype can only be answered when the other concepts to be acquired are considered. Consequently, as the interrelatedness of concepts increases, so should the categorization advantage of the caricature relative to the prototype.

EXPERIMENT 4

Experiment 4 tests the prediction that prototypes will be categorized relatively easily with instructions that bias subjects toward isolated concepts, whereas caricatures will be categorized relatively easily with instructions that bias subjects toward interrelated concepts. The experimental manipulation used is identical to the one used in Experiment 1. Thus, Experiment 4 is an attempt to provide

converging evidence for the results from Experiment 1, using a previously supported manipulation that affects concept interrelatedness and a new empirical indicator based on the ease of prototype/caricature categorization.

Experiments 1 and 2 indicated that nondiagnostic features were more likely to be used for categorization when few diagnostic features were altered from category prototypes. One explanation for this effect is that subjects were more likely to use a template-matching (an isolated representation) strategy for categorization when the overall similarity of the object to be categorized and its best-fitting category was high. In Experiment 4, this hypothesis was explored further, by varying the overall similarity of objects to prototypical category representations. On the majority of trials, displayed objects had particular nondiagnostic features. When these standard nondiagnostic features were present in a stimulus, it was predicted that subjects would more effectively use a template-matching strategy, because of the greater overlap between the item and the categories. The use of a template-matching strategy, in turn, would translate into greater facility with categorizing objects with prototypical, relative to caricatured, dimension values.

Also, in this experiment, four rather than two categories were used. One objection to Experiments 1–3 is that two-category conditions unnaturally bias subjects to use a special case of interrelated concepts in which one concept is

characterized simply as “whatever is not the other concept.” Creation of this type of “leftovers” category was discouraged in Experiment 4 by increasing the number of categories to learn.

Method

Materials. Sample materials are shown in Figure 5. The stimuli consisted of seven vertical bars joined together to form a histogram-like shape. Each bar assumed one of six different values (1.0 = shortest, 6.0 = tallest; each unit corresponded to 1.5 cm). The histograms belonged to one of four categories. Each category possessed one bar that was particularly long when compared with those for the other categories. These four bars, one for each category, were diagnostic, because if they had a height value above 1, then they provided information in favor of one and only one of the categories. The lengthened diagnostic bar for a category had a value of 2.0 on one quarter of the trials, 3.0 on one half of the trials, and 5.0 on one quarter of the trials. Thus, a value of 3.0 for a lengthened diagnostic bar was considered the prototypical value, because it occurred most frequently, and because its value was intermediate to the other possible values. A value of 5.0 for a diagnostic bar was considered a caricature, because it exaggerated the length of the bar that was particularly long for the category and it exaggerated the length in the direction opposite to that of the other categories’ bars.

The other three bars were nondiagnostic. At the beginning of an experiment, three randomly determined bar heights were generated. These were the *standard* values for the nondiagnostic bars. The bars were nondiagnostic because all four categories shared the same standard set of values for these bars. On one half of the trials, a histogram was presented with the standard values for all of the nondi-

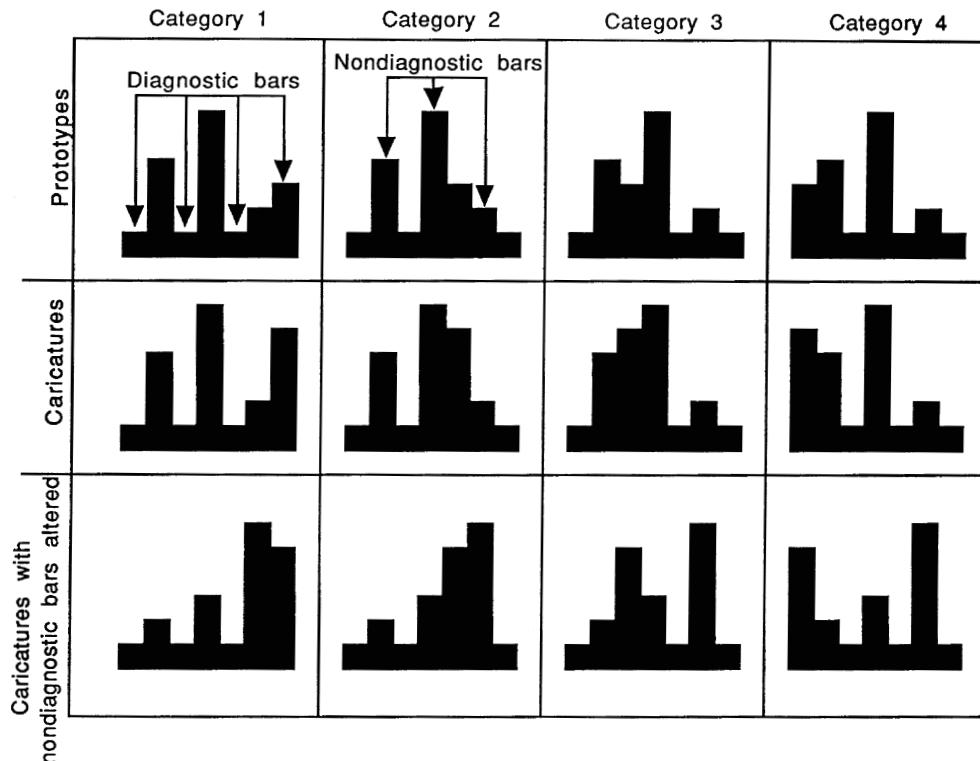


Figure 5. Sample stimuli from Experiment 4. Each of the four diagnostic bars, if lengthened, provides evidence in favor of one of the four categories. For example, Category 1 is suggested if the rightmost bar is lengthened. The standard heights of the nondiagnostic bars are the same for each of the four categories. Caricatures are created by further lengthening the height of a category’s diagnostic bar.

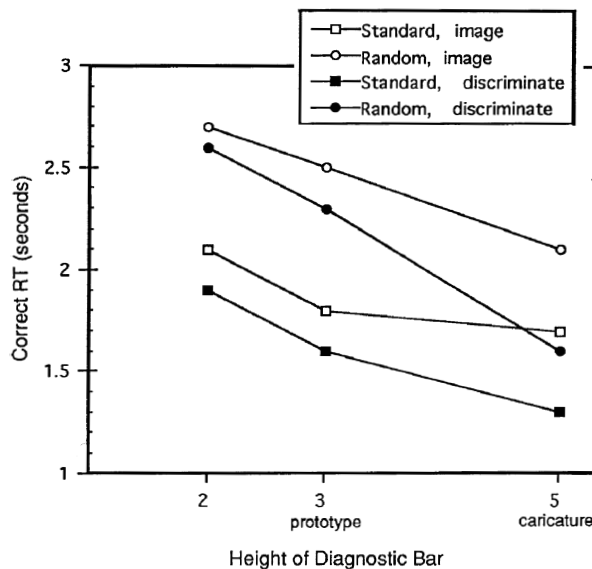


Figure 6. Results from Experiment 4. **Standard:** The nondiagnostic bars of the displayed item were given the standard values of the concept's prototype. **Random:** The nondiagnostic bars of the displayed item were assigned random values. Categorizing caricatures is generally faster than categorizing prototypes. This speed advantage is particularly pronounced when nondiagnostic bars are assigned random heights and when subjects are told to search for discriminating features.

agnostic bars. On the other half of the trials, new random heights were generated. Thus, subjects received a great deal of experience with one set of three nondiagnostic bar heights and received less experience with all other sets of nondiagnostic bar heights.

Procedure. Twenty-six undergraduates from Indiana University were divided into two instruction groups. The two sets of instructions corresponded to image and discriminate instructions used in Experiment 1. The subjects were reminded every 96 trials to use imagery or discriminating features to guide their categorization. At the beginning of the experiment, the subjects were instructed to make their categorization decisions as quickly as possible without sacrificing accuracy.

There were 384 trials in all. On each trial, a histogram appeared and subjects pressed one of four keys to indicate their proposed categorization for the object. The subjects then received feedback indicating whether their choice was correct and the correct response. All materials were presented on Macintosh SE computers. After all categorization trials were completed, subjects were asked to draw as good representations as possible of the four categories on 7×8 grid paper.

Results

The results of Experiment 4 are shown in Figure 6. Given that accuracy rates were above 93%, the dependent variable of greatest interest was response time for correct categorization. There was a general speed advantage for categorizing caricatures (height = 5, RT = 1.74 sec) relative to prototypes (height = 3, RT = 2.07 sec) [$F(1,24) = 14.7$, $MS_e = 0.47$, $p < .05$]. In addition, subjects were faster to categorize instances when the nondiagnostic bars were in their standard configuration (standard RT = 1.72, random RT = 2.27 sec) [$F(1,24) = 23.6$, $MS_e =$

0.49, $p < .05$]. Finally, there was a main effect of instructions, with discriminate subjects categorizing more quickly than image subjects (discriminate RT = 1.91 sec, image RT = 2.20 sec) [$F(1,24) = 11.1$, $MS_e = 0.26$, $p < .05$].

Figure 6 shows two interactions of interest—between height of lengthened diagnostic bar and instructions [$F(2,48) = 5.7$, $MS_e = 0.28$, $p < .05$], and between height and configuration of nondiagnostic bars [$F(2,48) = 6.5$, $MS_e = 0.19$, $p < .05$]. These interactions were still significant when only bar heights of 3 (prototype) and 5 (caricature) were considered [$F(1,24) = 5.0$, $MS_e = 0.35$, $p < .05$, and $F(1,24) = 5.2$, $MS_e = .40$, $p < .05$, respectively]. Both of these interactions are predicted by the isolated/interrelated framework. The first interaction indicates that the speed advantage of caricatures over prototypes was particularly pronounced when subjects were given discriminate (response time advantage = 452 msec) rather than image instructions (response time advantage = 187 msec). The second interaction indicates that the caricature advantage was greatest when items were presented with a random configuration of nondiagnostic features. One interpretation of this result is that when the overall stimulus configuration was similar to that of a familiar item, a process similar to template-matching transpired. Conversely, when the item, because of unfamiliar values for nondiagnostic features, was not similar to many previous items, rules of the form “third bar is relatively long” were used.

With increasing practice, the caricature advantage increased, as shown by a significant blocks (early vs. late) \times type of stimulus (prototype vs. caricature) interaction [$F(1,24) = 6.7$, $MS_e = 0.32$, $p < .05$]. For the first half of the 384 trials, correct categorizations took 2.6 and 2.5 sec for the prototypes and caricatures, respectively. For the second half, these numbers fell to 1.6 and 1.0, respectively.

The instructional manipulation also influenced subjects' pictorial representations of their concepts. Each of the four categories had one diagnostic bar that was lengthened. The actual modal height of this bar was 3.0 units. Image subjects estimated, on the average, the height of this bar to be 3.7, whereas discriminate subjects estimated this height to be 4.3 [$t(24) = 3.3$, $p < .05$]. The heights of the three other diagnostic bars were 1.0. Image subjects estimated these bars to be 1.2, whereas discriminate subjects gave an average estimate of 1.4, a nonsignificant difference [$t(24) = 1.0$, $p > .1$].

Discussion

The nature of the prototype/caricature advantage in categorization was influenced by both task and stimulus factors. When subjects looked for discriminating features, there was evidence of a strong advantage for categorizing caricatures relative to prototypes. This caricature advantage was also found when the item to be categorized did not share nondiagnostic features with previously categorized items. The caricature advantage was reduced when either of these factors was altered—that is, when subjects were given image instructions, or when nondiagnostic fea-

tures were preserved. These results are consistent with those of the first three experiments and support the validity of a second indicator of concept interrelatedness.

EXPERIMENT 5

Experiment 5 was conducted to obtain one final converging source of evidence, using the task manipulation of Experiment 2 and the indicator of interrelatedness developed in Experiment 4. Category instances were alternated either frequently or infrequently, and the effect of this manipulation on caricature/prototype categorization was observed.

Method

The materials from Experiment 4 were used. The procedure from Experiment 4 was used, with a few exceptions. Subjects were given no instructions on the appropriate categorization strategy to use. Twenty-four Indiana University undergraduates were split into two groups. For one group of subjects, the *frequent alternation* group, the probability of presenting an item from the same category as that of the previous item was .07, and the probability of presenting an item from one of the three other categories was .31. For the *infrequent alternation* group, the probability of presenting an item from the same category as that of the previous item was .31 and the probability of presenting an item from one of the three other categories was .23. If subjects used an optimal guessing strategy and had no knowledge of an item's category, the two conditions were equated for probability of correct categorization (.31). Compared with the alternation manipulation in Experiment 2, the overall frequency of alternation was higher in Experiment 4 because four, rather than two, categories were used.

Results and Discussion

The results for Experiment 5 are shown in Figure 7. As with Experiment 4, caricatures were generally categorized more quickly than prototypes (prototype RT = 1.78, caricature RT = 1.51) [$F(1,22) = 13.2, MS_e = 0.22, p < .05$]. Subjects made faster categorizations when nondiagnostic features were given familiar values (as with Experiment 4) [$F(1,22) = 12.9, MS_e = 0.32, p < .05$] and when concepts were alternated infrequently (as with Experiment 2) [$F(1,22) = 9.8, MS_e = 0.34, p < .05$]. Categorization accuracy was 94% for the infrequent alternation group and 92% for the frequent alternation group [$F(1,22) = 1.1, MS_e = 0.58, p > .1$].

There were also two interactions involving the height of the lengthened diagnostic bar. First, the speed advantage for caricature over prototype categorization was greater when concepts were alternated frequently rather than infrequently [$F(2,44) = 5.4, MS_e = .18, p < .05$]. Second, the speed advantage for caricature over prototype categorization was greater when nondiagnostic features were given unfamiliar rather than familiar values [$F(2,44) = 3.87, MS_e = .13, p < .05$]. These results are consistent with the predictions of the isolated/interrelated framework. Frequent alternation of categories was expected to yield interrelated concepts, because instances from different concepts would more likely be compared to each other. By this hypothesis, the presence of familiar nondiagnostic bar values increased the likelihood of

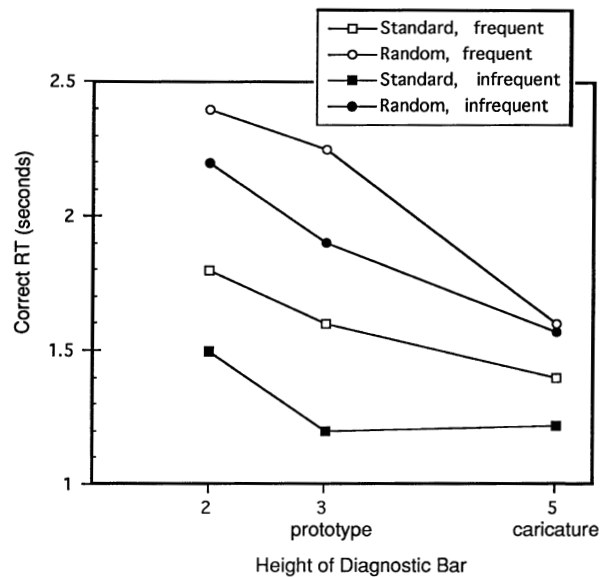


Figure 7. Results from Experiment 5. Standard: The nondiagnostic bars of the displayed item were given the standard values of the concept's prototype. Random: The nondiagnostic bars of the displayed item were assigned random values. A strong caricature advantage is found when nondiagnostic bars are assigned random heights and when the categories are alternated frequently.

using isolated concepts because it facilitated an overall template match rather than a selective use of discriminating features.

As with Experiment 4, the caricature advantage increased with practice, as evidenced by a two-way interaction between block (early vs. late) and type of stimulus (caricature vs. prototype) on response time [$F(1,22) = 7.1, MS_e = 0.28, p < .05$]. For the first half of the 384 trials, correct categorizations took 2.0 and 1.9 sec for the prototypes and caricatures, respectively. For the second half, these numbers fell to 1.5 and 1.1, respectively.

GENERAL DISCUSSION

The five experiments were performed to investigate predictions made by the claim that concepts vary in the degree to which they are influenced by other simultaneously acquired concepts. The suggested continuum between isolated and interrelated concepts motivated the development of new task manipulations and measurements. The degree of interrelatedness was quantified in terms of the influence of nondiagnostic features on categorization, the prevalence of nondiagnostic features in subjects' drawings, the relative advantage to categorizing prototypes relative to caricatures, and the degree of caricaturization in drawings. These four measures were in close agreement in locating concepts on the continuum of isolation/interrelation.

Reciprocally, the manipulations are in close accord in their influence of the measures of isolation. Isolated concepts are promoted by giving subjects imagery instructions, by alternating concepts infrequently, by assigning

unrelated labels to concepts, by testing subjects early in practice, and by presenting familiar or less distorted instances. Predictions for the first three manipulations are established by simple analysis of the manipulations. If subjects are asked to create separate images for each concept, if instances from different concepts are not often presented together, and if labels are presented that do not relate the concepts, then we expect that the concepts will become relatively isolated from each other. The final two manipulations (increasing the amount of training produces relatively interrelated concepts, and decreasing the similarity between an object and its category produces relatively interrelated concepts) have weaker a priori associations with concept isolation, but consistently cohere with the other three manipulations. The empirical results, from all four measures of interrelatedness, argue that concepts are increasingly influenced by each other as concept learning continues, and as increasingly distorted concepts are displayed.

Other Tests of the Interrelated/Isolated Framework

Experiment 3 introduced a manipulation that created asymmetrically defined concepts. Asymmetries were also found from manipulations of frequency and presentation order. If one concept is presented more frequently or earlier than another concept, it is predicted to be relatively isolated. The less frequent, later concept will tend to be influenced by the frequent, earlier concept. These predictions have been tested with nondiagnostic feature use as an indicator of concept isolation (Goldstone, 1996). Nondiagnostic features were more likely to be used for categorization, and depicted in concept representations, for categories that were presented on 80%, rather than 20%, of trials. Similarly, nondiagnostic features had more of an influence on concept depictions and categorizations for categories that were presented in the first 200, rather than the second 200, trials. Unfortunately, presentation order was confounded with practice effects in this experiment.

In addition to Experiment 3, a second labeling manipulation was tested (Goldstone, 1996). Stick-figure prototypes similar to those in Figure 1 were constructed for three categories. These three prototypes can be called A, B, and C. Every pair of category prototypes (A and B, A and C, and B and C) had two line segments in common. Although the prototypes were logically equivalent, one randomly picked prototype was labeled "Art from a Venusian Colony," and the other two prototypes were labeled "Art from Martian Colony 1" and "Art from Martian Colony 2." It was predicted that the two Martian categories would be more interrelated to each other than they would be to the Venusian category, because the labels made them more likely to be contrasted from each other than they would otherwise be. Consistent with this prediction, for the category labeled "Art from Martian Colony 1," line segments that distinguished the category from art by Martian colony 2 were more influential than the line segments that distinguished the prototype from art by the Venusian colony. Thus, when labeling caused

two categories to be compared and contrasted, features that discriminated between the categories were selectively highlighted.

One of the main benefits of the isolated/interrelated framework is that it provides an integrated account for results from many different paradigms. For behavioral indicators, the framework links responding to caricatures accurately, being highly influenced by nondiagnostic features, and being highly influenced by intercategory similarities (Goldstone, 1996). The framework also links all of the experimental manipulations shown in Table 1. However, it is possible to explain particular experimental results without hypothesizing a continuum between isolated and interrelated concepts. In fact, a full account of the reported experiments will require the development of particular processing accounts for the experimental manipulations tested. Still, the framework is a valuable first step, because it predicts the observed commonalities between apparently different tasks and indicators. For many of the manipulations, particularly those that involve labeling, we are a long way from describing adequate process models, but we can still give a general characterization of the tasks as biasing subjects toward isolated or interrelated concept representations.

Similarly, the ramifications of the present experiments for natural semantic categories are not yet clear. While the present experiments were focused on task manipulations that influenced concept interrelatedness, there may be systematic differences in the degree of interrelatedness of different classes of natural concepts. The first potential type of group difference is that abstract concepts may be more interrelated than concrete, perceptual objects. The basis for this prediction is that the most readily available methods for representing isolated concepts, including feature detectors and templates, seem to be tied to perceptual features. However, nothing in the present results or discussion requires that interrelated concepts be conceptual or isolated concepts be perceptual. In fact, all of the concepts acquired in Experiments 1–5 were predominantly visual, and one of the sources of evidence for interrelatedness was based on the perceptual caricaturization of concepts.

A second potential difference is that artifact categories may be more interrelated than natural kind categories. Barr and Caplan (1987) find that artifacts have many extrinsic features associated with them (a feature of *hammer* is that it is used to hit nails), whereas natural kinds have a preponderance of internal, intrinsic features (a feature of *robin* is that it has wings). Third, real-world concepts may be more interrelated than fictional concepts. Potts et al.'s (1989) subjects read a story about a novel bird and were told either that the bird was fictional or that it truly existed. Subjects' concept of the bird was more isolated from the rest of subjects' knowledge (as measured by the time required to answer questions about the bird that are presented in a context not related to the story) when they believed the story to be fictional rather than real.

There is a difference between concepts that are *currently* dependent on other concepts for their characteriza-

tion, and concepts that are dependent on other concepts during their *development* but not in their current state. The measures used in the present experiments did not distinguish between these options. As such, evidence for interrelatedness in the experiments will only be taken as evidence that the concepts influenced each other's characterization at some time, and may still actively influence each other. One potential way to obtain evidence for active dependencies between concepts is to systematically alter one of the concepts and observe changes in the other. After two categories are learned, new information can be provided about one of the concepts. If the other concept has an "active" link to the concept or if the two concepts share representational units, the other concept's representation should change.

Potential Categorization Models

The primary implication of the results for psychological models is that a full model of categorization should be able to have categories exert varying degrees of influence on each other. In order to account for the empirical results, models should be able to predict both main effects (diagnostic features are more influential than nondiagnostic features, nondiagnostic features exert some influence, and caricatures are categorized faster than prototypes) and the interactions between these effects and task manipulations. In this section, the ability of existing categorization models to account for the main effects and interactions is assessed.

Due to their processing principles, prototype (Posner & Keele, 1968; Reed, 1972; Rosch, 1975) and exemplar (Medin & Schaffer, 1978; Nosofsky, 1986) models can accommodate results that have been taken to indicate relatively interrelated concepts such as a caricature advantage (Nosofsky, 1991; see also the classic "peak shift" effect handled by Spence, 1936) and selective influence of diagnostic features (Rosch, Simpson, & Miller, 1976). Categorization models can accommodate a differential influence of nondiagnostic and diagnostic features on categorization accuracy, even if concept representations in the models are not influenced by other concepts. For example, in the context model of Medin and Schaffer (1978), the probability of an item i being placed in a category n is equal to

$$P(C_n | i) = \frac{\sum_{j \in C_n} S_{ij}}{\sum_k \sum_{j \in C_k} S_{ij}},$$

where S_{ij} is the similarity of items j and i . Thus, the probability of placing an item in a category is equal to the summed similarity of the item to every example in the category n divided by the summed similarity of the item to all examples from all categories. If we further assume, as Medin and Schaffer do, that S_{ij} is determined by multiplying dimensional proximities (if two items have the same value on a dimension, then a value of 1 is assigned; otherwise, a value between 0 and 1 is assigned), it is ev-

ident that nondiagnostic features are not expected to influence categorization accuracy. For example, consider two categories, with Category A consisting of {red square, red circle, red diamond} and Category B consisting of {red hexagon, red triangle, red pentagon}. Assume that a color mismatch results in a dimensional proximity of d , and any shape mismatch results in a dimensional proximity of e . Presenting a red square results in

$$\begin{aligned} P(\text{category A} | \{\text{red square}\}) \\ = \frac{1 + e + e}{(1 + e + e) + (3e)} = \frac{2e + 1}{5e + 1}, \end{aligned}$$

whereas presenting a blue square results in

$$\begin{aligned} P(\text{category A} | \{\text{blue square}\}) \\ = \frac{d + 2de}{(d + 2de) + (3de)} = \frac{2de + d}{5de + d} = \frac{2e + 1}{5e + 1}. \end{aligned}$$

Thus, whether the item to be categorized possesses the nondiagnostic feature (red) that is characteristic of both categories should not influence categorization. Goldstone (1993b) contains a formal demonstration of this conclusion for the particular materials and the methods used in Experiment 1.

The context model and the Hull–Spence model show that it is possible to develop models that account for evidence of interrelated concepts without positing that concepts influence each other's representations. Instead, these model predict advantages for caricatures and diagnostic features by decisional rules that use evidence for one response *relative* to evidence for another response. In the case of the context model, the nondiagnostic and diagnostic features have different influences because of the inherently relative categorization rule (a ratio rule). The results from Experiments 1–5 suggest two problems for this purely decisional approach. First, results from tasks where subjects provide depictions of their category knowledge are not naturally accommodated. In all five experiments, subjects were asked to draw their best possible representation of the acquired concepts. Results showed that diagnostic features were generally more likely to be correctly drawn than nondiagnostic features, and that there was a bias to draw caricatured, as opposed to prototypical, dimension values. Importantly, the draw-a-concept task requires an absolute, not a relative, judgment (Busemeyer & Myung, 1988). Deciding what category to place an object in involves consideration of all of the learned categories; the judgment is an inherently relative one. By contrast, there is nothing in the instructions to draw an accurate representation of a typical concept member that should dispose subjects to consider other concepts. The empirical result that caricatures and diagnostic features dominate drawings suggests that the characterization of an acquired concept is influenced by other concepts, even when the concept's characterization is probed with tasks that are not relational.

The second difficulty for these particular models is in accommodating the influences of the task manipula-

tions. It is not sufficient to develop a model that can predict a caricature or prototype advantage, or predict that nondiagnostic features either will or will not influence categorization accuracy. An account is also needed that describes the conditions under which each of the results is likely to be found. For example, nondiagnostic features are more likely to exert an influence under conditions that have been grouped together under the label “relatively isolated.” These conditions include imagery instructions, infrequent category alternation, and unrelated labeling. If we assume that the degree of concept interrelatedness is variable, and if we make reasonable assumptions about how particular tasks affect interrelatedness, the observed influences of the task manipulations are predicted. If we assume only isolated manipulations, no systematic explanation is given for the task manipulations’ effects. Thus, we gain confidence in the isolated/interrelated continuum because its task manipulation predictions are borne out empirically. Although there may be alternative explanations to account for the effect of some of the task manipulations separately, the isolated/interrelated framework provides a single coherent account for the entire set of manipulations.

Future research with exemplar or prototype models may generate parameters that are naturally associated with the task manipulations used in Experiments 1–5 and that correctly predict when caricatures are well categorized and nondiagnostic features are influential. Until this happens, it seems that the most parsimonious synthesis of the results is that experimental manipulations differentially affected how interrelated concepts were. This interpretation naturally captures commonalities of the experimental manipulations, captures commonalities between the two experimental measures (prototype/caricatures categorizations and nondiagnostic/diagnostic feature influence), and predicts the results from measures of interrelatedness that involve category reconstructions rather than categorization judgments. The results suggest that complete models of categorization should incorporate concept characterizations, through representation and process, that can vary in their degree of interrelatedness.

A Connectionist Approach to Isolated/Interrelated Concepts

A connectionist model, RECON, has been developed to provide a qualitative account of the experiments’ results. The purpose of the model is to provide a demonstration of how a single mechanism can yield both isolated and interrelated concepts, rather than to provide detailed quantitative fits (such as those provided by Kruschke, 1992; Nosofsky, 1986). Essentially, RECON is a two-layer recurrent network. One layer of units represents the input dimensions, and one layer represents the categories being learned. All units in RECON are connected to each other by weighted connections. The most important elements of RECON are its recurrent connections between category units and from category units to input units. These connections provide a mechanism for categories to influence each other and for a featural description to be produced

when a category label is given. By varying a single parameter, the degree of influence of category units on each other, varying degrees of concept interrelatedness are obtained.

In feed-forward connectionist systems, activation flows from input to hidden to output units, and the weights that are learned regulate this unidirectional flow. Typically, input units encode the featural or dimensional representation of the object to be categorized, and each output unit signifies a category. In RECON, activation flows *between* output/category units, from output to input units, and between input units, in addition to the standard feed-forward flow. Recurrent activation passing is a feature of several connectionist models, including McClelland and Rumelhart’s (1981) interactive activation model and Grossberg’s ART system (Carpenter & Grossberg, 1991).

The details of RECON’s spread of activation are found in Appendix. Learning in RECON consists of two stages: recurrent activation passing, followed by a weight adjustment procedure following Rumelhart, McClelland, and the PDP Research Group (1986). In the first stage, units that represent input dimensions and categories spread activation between each other. The spread of activation is modulated by learned connection weights. Activation is spread for a number of cycles determined by the parameter *cycles*. In the second stage, weights are adjusted so that the units that have similar activations after the allotted number of activation-passing cycles become more strongly connected. When RECON is tested after training, the recurrent flow of activation between units occurs when a pattern is presented, and in this way, concepts are influenced by each other and influence input representations. The concept-to-input influence provides a way to account for results that suggest that categorization training can change low-level perceptual sensitivity to objects (Goldstone, 1994a).

Experimental manipulations of interrelatedness are modeled by manipulating category-to-category weights. Category-to-category weights are clamped to particular values. The assumption is that tasks that yield highly interrelated concepts are modeled with category-to-category weights with relatively large absolute values. If category-to-category weights are small, the activation of one category will not depend on the other categories’ activations very much. Increasing the strength of connection weights between categories is expected to produce performance that is similar to that of subjects who were given task manipulations that yielded interrelated concepts.

Nondiagnostic and diagnostic features. Increasing the strength of category-to-category weights (by making the weights increasingly negative) yields a larger disparity between the influence of diagnostic and nondiagnostic features. One of the simplest tests of the influence of nondiagnostic and diagnostic features is to provide RECON with the two patterns: 1 0 1 0 1 0, and 1 0 0 1 0 1, as shown in Figure 8. These two patterns are presented to RECON 100 times each, in random order. In these patterns, the first two values code for one binary dimension (“1 0” can be interpreted as “red = yes, blue = no”), the third

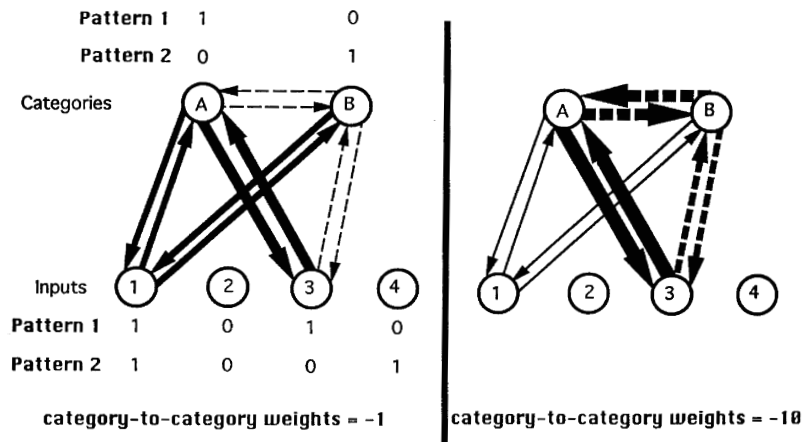


Figure 8. RECON networks that model the influence of nondiagnostic and diagnostic features on categorization under two different values for category-to-category weights. Dashed and solid lines represent negative and positive connection weights, respectively. The thickness of a line represents the absolute magnitude of a connection. Diagnostic units (e.g., input unit 3) are relatively more influential than nondiagnostic units (e.g., input unit 1) as the magnitude of category-to-category weights increases.

and fourth values code for a second binary dimension (“1 0” = large, “0 1” = small), and the last two values code the category (“1 0” can be interpreted as “Category A = yes, Category B = no”). By using two units to code dimensions and categories, conceptual problems with +1/−1 units are avoided (e.g., how are absent features represented—as −1 or 0?) and several potential relations between concepts can be modeled (Categories A and B may be negatively connected if mutually exclusive, or positively connected if hierarchically or associatively related).

The pattern “1 0” on the first two units is nondiagnostic for categorization. Both the A and B category items have the same values for this dimension. This pattern has high category validity though, because all items possess this pattern. The second dimension is diagnostic. If the pattern for the second dimension is “1 0,” the category pattern is “1 0.” If the pattern is “0 1,” the category pattern is “0 1.”

Figure 8 shows RECON’s architecture for the problem, and the learned weights under two values of *category-to-category weights*. All weights are directional, with solid and dashed lines indicating positive and negative weights, respectively. The thickness of a connection reflects the magnitude of the weight. Only some of the connections are shown; the actual network is fully connected. Three sources of evidence point to an influence of category-to-category weights on the importance of diagnostic/nondiagnostic features.

First, the difference between the dimension-to-category weights for nondiagnostic and diagnostic features is greater as category-to-category weights become increasingly negative. When category-to-category weights are −1, the connection strength from nondiagnostic and diagnostic features to the categories is roughly equal after training on the two patterns. When categories exert more

influence on each other, achieved by increasing category-to-category weights to −10, the connection from diagnostic features to categories is much stronger than the connection from nondiagnostic features to categories.

Second, nondiagnostic features have decreasing influence on categorization accuracy with increasingly negative category-to-category weights. Test trials consist of the patterns “0 0 1 0 0 0” (nondiagnostic feature removed) and “1 0 1 0 0 0” (nondiagnostic feature present). During testing, RECON is given one of the patterns and earning is disabled. The likelihood of giving the correct “1 0” category response is substantially reduced when the nondiagnostic feature is removed, but only when the categories do not have strongly negative connecting weights. If category-to-category weights are strongly negative, the two patterns result in almost identical categorization accuracies.

Third, relative to diagnostic features, nondiagnostic features are less prominently figured in concept representations when category-to-category weights are strongly negative. Because of the category-to-dimensions connections in RECON, we can observe the resulting spread of activation from the pattern “0 0 0 0 1 0.” This is tantamount to telling RECON that an item belongs to Category A, and asking RECON what it looks like. As such, it provides a method for modeling the draw-a-concept tasks of Experiments 1–5. Although both nondiagnostic and diagnostic units are activated by activating a category node, diagnostic units become much more activated with increasingly negative category-to-category weights. If category-to-category weights are only −1, nondiagnostic and diagnostic features are almost equally strongly activated by “0 0 0 0 1 0.”

All three of these effects are attributable to a single cause. If the connection between categories is strongly negative (the concepts are relatively interrelated), then, when one category unit begins to be activated by the input,

it will inhibit the other category unit. If the inhibition is sufficiently strong, the less activated unit will develop a negative activation. If this occurs during the recurrent activation passing stage, then, during the learning stage, there will be a negative correlation between the present nondiagnostic features' activation and the inhibited category's activation. According to the Hebbian learning rule, this negative correlation will result in a decrease in the input-to-category connection weight. This learned weight change will counteract the increase in the weight that occurs when one category dominates over the other category. As a result, the nondiagnostic feature will not be strongly associated with either category.

If the connection between categories is not strong (the concepts are relatively isolated), no category unit will be strongly inhibited by other categories, and there will be positive correlations between the present nondiagnostic features and both category units. As a result, the association between the nondiagnostic features and both categories will increase.

In short, when category units strongly inhibit each other, diagnostic features become much more important

for categorization than do nondiagnostic features. Strongly negative category-to-category weights result in competition between the categories. After activation has been passed for several cycles, only one category will have a strong positive activation. In these circumstances, features that do not distinguish between categories will not be highly associated with either category. This effect has been replicated in a simulation that involved 20 input nodes and 600 randomly generated distortions of two category prototypes constructed in the same manner as that in which they were constructed in Experiments 1-3.

Prototype and caricature categorizations. The categorization advantage of caricatures over prototypes is increased by making category-to-category weights increasingly negative. This is apparent from an analysis of categorization performance, concept production, and connection strengths. One of the simplest sets of materials for exploring caricature/prototype categorizations is shown at the top of Figure 9. Six patterns are placed in two categories. Six units are used to represent a single underlying dimension, and each example is unidimensional. The six patterns shown in Figure 9 are randomly presented to

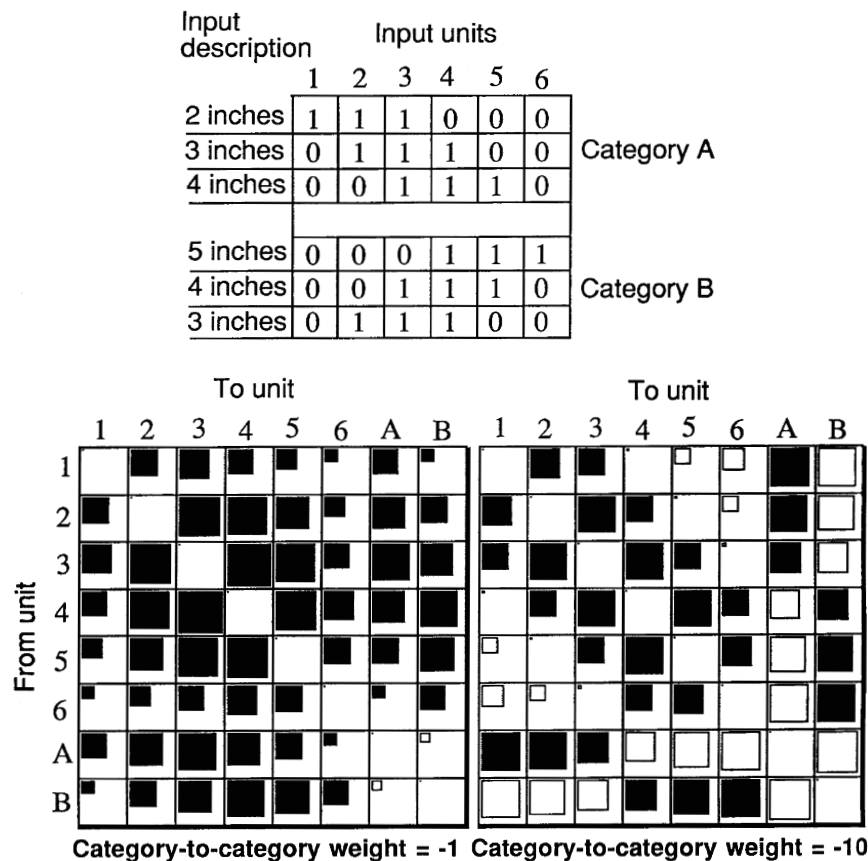


Figure 9. A simulation of caricature/prototype advantages in RECON. The top of the figure depicts the 6 training trials presented to RECON. Category A consists of inputs representing values of 2, 3, and 4 on a dimension. Category B consists of inputs representing values of 3, 4, and 5. If weights between categories have small absolute magnitudes, the prototypical values are most strongly associated with the categories. If categories exert more influence on each other, the caricature values are associated with categories almost as strongly as are prototypical values.

RECON 100 times each during learning. The dimension value of an item is represented by the heightened activity of units that are linearly ordered (for another example of "place coding" in connectionist architectures, see McClelland & Jenkins, 1991). The three input patterns to Category A can be thought of as coding for 2, 3, and 4 in., whereas the three input patterns to Category B are 5, 4, and 3 in. The prototype for Category A is 3 in., and its caricature is 2 in. As before, Category A is represented by "0 1" on two additional units, and Category B is represented by "1 0."

As is evident in Figure 9, the connection weights between input and category units depends on the strength of category-to-category weights. When category-to-category weights are -1 , the input unit with the strongest connection to Category A is the one that codes for 3 in. When category-to-category weights are -10 , the input that codes for 1 in. is most strongly connected to Category A, and the unit that codes for 4 in. is negatively connected to Category A. Thus, as categories increasingly inhibit each other, the category becomes increasingly associated with its caricature rather than its prototype.

A similar trend from strong prototype to strong caricature connections exists for the weights that connect categories to inputs. Because of these reciprocal connections, the draw-a-concept results are accommodated. When category-to-category weights are -1 and the representation "0 0 0 0 0 1 0" is given as input (i.e., only the category name is provided), the inputs attain values in which the prototype is much more active than the caricature (0.14 vs. 0.01). When category-category weights are -10 , the caricature is more active than the prototype (0.19 and 0.13, respectively).

Finally, when categorization judgments are requested, the expected interaction between *cycles* and prototype/caricature advantage is found. When category-to-category weights are -1 , Category A is more likely evoked by "0 0 1 0 0 0 0" (i.e., only the prototype unit activated) than by "0 1 0 0 0 0 0" (caricature). When category-to-category weights are -10 , there is a categorization advantage for the caricature.

RECON, rather than being a competitor to current connectionist models of categorization (Gluck & Bower, 1988; Kruschke, 1992), is properly viewed as a method for augmenting these models. Architectural aspects of these models, such as learned selective attention to diagnostic dimensions (Kruschke, 1992), would have to be incorporated into RECON for it to be a full model of categorization. RECON's value is that it furnishes possible methods for accommodating task manipulations, yielding concepts that occupy various positions along the isolated/interrelated continuum. RECON cannot model all of the specific results from the experiments, but it does show how a single model can develop varying degrees of intercategory influence. Furthermore, it provides an account for the empirical correlation found between dependent measures. In RECON, the same parameter manipulation

that produces a caricature advantage also produces a relatively strong influence of diagnostic features.

CONCLUSION

The five experiments support the postulation of a continuum between completely isolated and completely interrelated concepts. It is experimentally possible to manipulate the degree of isolation of a concept, as measured by a variety of converging operations. A connectionist framework accounts for the results by assuming recurrent connections among concept units and from concept units to input units. Interrelated concepts are modeled by allowing concept units to exert a relatively large influence on each other. This line of work offers the promise of providing a bridge between two communities that study concepts. On the one hand, as many psychologists who study language argue, concepts are intricately connected to each other, and these connections influence the internal representations of concepts. On the other hand, as exemplified by models of object recognition, concepts also seem to be directly accessed for the purpose of recognition, using representations that are partially independent of other concepts. In fact, concepts may typically be located at intermediate positions along a continuum of interrelatedness such that they gain their meaning in part from relations to other concepts and in part from external grounding.

REFERENCES

- ANDERSON, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409-429.
- ASHBY, F. G., & GOTT, R. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Human Perception & Performance*, *14*, 33-53.
- BARR, R. A., & CAPLAN, L. J. (1987). Category representations and their implications for category structure. *Memory & Cognition*, *15*, 397-418.
- BARSALOU, L. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A. C. Collins, S. E. Gathercole, M. A. Conway, & P. E. M. Morris (Eds.), *Theories of memory* (pp. 29-101). Hillsdale, NJ: Erlbaum.
- BIEDERMAN, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115-147.
- BROWNELL, H. H., & CARAMAZZA, A. (1978). Categorizing with overlapping categories. *Memory & Cognition*, *6*, 481-490.
- BRUCE, C., DESIMONE, R., & GROSS, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology*, *46*, 369-384.
- BUSEMEYER, J. R., & MYUNG, J. (1988). A new method for investigating prototype learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *14*, 3-11.
- CAPLAN, L. J., & BARR, R. A. (1991). The effects of feature necessity and extrinsicity on gradedness of category membership and class inclusion relations. *British Journal of Psychology*, *82*, 427-440.
- CARPENTER, G. A., & GROSSBERG, S. (1991). *Pattern recognition by self-organizing neural networks*. Cambridge, MA: MIT Press.
- CLARK, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language*, *17*, 417-431.
- COLLINS, A. M., & QUILLIAN, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning & Verbal Behavior*, *8*, 240-247.

- CORTER, J. E., & GLUCK, M. A. (1992). Explaining basic categories: Feature predictability and information. *Psychological Bulletin*, **111**, 291-303.
- FODOR, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press, Bradford Books.
- GARNER, W. R., HAKE, H. W., & ERIKSEN, C. W. (1956). Operationism and the concept of perception. *Psychological Review*, **63**, 149-159.
- GLUCK, M. A., & BOWER, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, **117**, 227-247.
- GOLDSTONE, R. L. (1991). Feature diagnosticity as a tool for investigating positively and negatively defined concepts. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* (pp. 263-268). Hillsdale, NJ: Erlbaum.
- GOLDSTONE, R. L. (1993a). Feature distribution and biased estimation of visual displays. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 564-579.
- GOLDSTONE, R. L. (1993b). *Positively and negatively defined concepts* (Tech. Rep. No. 88). Bloomington: Indiana University, Cognitive Science Program.
- GOLDSTONE, R. L. (1994a). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, **123**, 178-200.
- GOLDSTONE, R. L. (1994b). The role of similarity in categorization: Providing a groundwork. *Cognition*, **52**, 125-157.
- GOLDSTONE, R. L. (1996). *The influences of between- and within-category similarity on isolated and interrelated concepts*. Manuscript in preparation.
- GRANDY, R. E. (1987). In defense of semantic fields. In E. Le Pore (Ed.), *New directions in semantics* (pp. 261-280). New York: Academic Press.
- HARNAD, S. (1990). The symbol grounding problem. *Physica D*, **42**, 335-346.
- HOCK, H. S., TROMLEY, C., & POLMANN, L. (1988). Perceptual units in the acquisition of visual categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 75-84.
- HUBEL, D. H., & WIESEL, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Neurophysiology*, **195**, 215-243.
- HUNT, R. R., & MCDANIEL, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory & Language*, **32**, 421-445.
- JOHNSON-LAIRD, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.
- KRUEGER, J., & ROTHBART, M. (1990). Contrast and accentuation effects in category learning. *Journal of Personality & Social Psychology*, **59**, 651-663.
- KRUSCHKE, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.
- LAKOFF, G. (1987). *Women, fire and dangerous things: What categories tell us about the nature of thought*. Chicago: University of Chicago Press.
- LEHRER, A., & KITTAY, E. F. (1992). *Frames, fields and contrasts*. Hillsdale, NJ: Erlbaum.
- MALSBURG, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, **14**, 85-100.
- MARKMAN, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, **14**, 57-77.
- MARTIN, J. D., & BILLMAN, D. O. (1994). Acquiring and combining overlapping concepts. *Machine Learning*, **16**, 121-155.
- MCCLELLAND, J. L., & JENKINS, E. (1991). Nature, nurture, and connections: Implications of connectionist models for cognitive development. In K. VanLehn (Ed.), *Architectures for intelligence* (pp. 41-73). Hillsdale, NJ: Erlbaum.
- MCCLELLAND, J. L., & RUMELHART, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, **88**, 375-407.
- MCMANARA, T. P., & MILLER, D. L. (1989). Attributes of theories of meaning. *Psychological Bulletin*, **106**, 355-376.
- MEDIN, D. L., & EDELSON, S. M. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General*, **117**, 68-85.
- MEDIN, D. L., & SCHAFFER, M. M. (1978). A context theory of classification learning. *Psychological Review*, **85**, 207-238.
- MEDIN, D. L., WATTENMAKER, W. D., & MICHALSKI, R. S. (1987). Constraints in inductive learning: An experimental study comparing human and machine performance. *Cognitive Science*, **11**, 319-359.
- NOSOFSKY, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.
- NOSOFSKY, R. M. (1991). Typicality in logically defined categories: Exemplar-similarity versus rule instantiation. *Memory & Cognition*, **19**, 131-150.
- OSHERSON, D. N., SMITH, E. E., WILKIE, O., LOPEZ, A., & SHAFIR, E. (1990). Category-based induction. *Psychological Review*, **97**, 185-200.
- PALMER, S. E. (1978). Structural aspects of visual similarity. *Memory & Cognition*, **6**, 91-97.
- POSNER, M. I., & KEELE, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, **77**, 353-363.
- POTTS, G. R., ST. JOHN, M. F., & KIRSON, D. (1989). Incorporating new information into existing world knowledge. *Cognitive Psychology*, **21**, 303-333.
- REED, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, **3**, 382-407.
- RHODES, G., BRENNAN, S., & CAREY, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology*, **19**, 473-497.
- ROSCH, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: Human Perception & Performance*, **1**, 303-322.
- ROSCH, E., & MERVIS, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, **7**, 573-605.
- ROSCH, E., SIMPSON, C., & MILLER, R. S. (1976). Structured bases of typicality effects. *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 491-502.
- RUMELHART, D. E., & MCCLELLAND, J. L., & THE PDP RESEARCH GROUP (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1*. Cambridge, MA: MIT Press.
- RUMELHART, D. E., & ZIPSER, D. (1985). Feature discovery by competitive learning. *Cognitive Science*, **9**, 75-112.
- SAUSSURE, F. (1959). *Course in general linguistics*. New York: McGraw-Hill. (Original work published 1915)
- SCHYNS, P., GOLDSTONE, R. L., & THIBAUT, J. (in press). Development of features in object concepts. *Behavioral & Brain Sciences*.
- SPENCE, K. W. (1936). The nature of discrimination learning in animals. *Psychological Review*, **43**, 427-429.
- ULLMAN, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, **32**, 193-254.
- VAN WALLENDAL, L. R., & HASTIE, R. (1990). Tracing the footsteps of Sherlock Holmes: Cognitive representations of hypothesis testing. *Memory & Cognition*, **18**, 240-250.
- WAXMAN, S. R., CHAMBERS, D. W., YNTEMA, D. B., & GELMAN, R. (1989). Complementary versus contrastive classification in preschool children. *Journal of Experimental Child Psychology*, **48**, 410-422.
- WAXMAN, S. R., & SENGHAUS, A. (1992). Relations among word meanings in early lexical development. *Developmental Psychology*, **28**, 862-873.
- WHITMAN, J. R., & GARNER, W. R. (1963). Concept learning as a function of the form of internal structure. *Journal of Verbal Learning & Verbal Behavior*, **2**, 195-202.
- WINSTON, M. E., CHAFFIN, R., & HERRMANN, D. (1987). A taxonomy of part-whole relations. *Cognitive Science*, **11**, 417-444.
- WISNIEWSKI, E. J., & MEDIN, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science*, **18**, 221-281.

NOTE

1. It is important to draw a distinction between word meaning and concepts. That there is no equivalent of *mutton* in French does not mean that French speakers lack this concept. Saussure was primarily interested in word meanings, but he also extended his argument for completely interrelated meanings to concepts in general.

APPENDIX
Formal Details of RECON

The number of cycles of activation passing was arbitrarily set at 15 for all simulations. Activation for both input dimensions and categories is in the range -1 to $+1$. Connection weights between input and category units are initially set to zero, and weights between category units are clamped to a constant value. When processing a new trial, activation is first spread between all units (units are not connected to themselves). The input activation to a unit j , net_j , is

$$net_j = extr(e_j) + intr \sum_i w_{ji} a_i,$$

where e_j is the externally provided activation of unit j , w_{ji} is the weight of the connection from unit i to j , the internal and external strengths (*extr* and *intr*) are both set to 0.15, and a_i is the current activation of unit i (0.15 was selected because of its use in simulations by Rumelhart et al., 1986). The change of activation of unit j can now be expressed as

$$\Delta a_j = \begin{cases} net_j(\max - a_j) & \text{if } net_j > 0 \\ net_j(a_j - \min) & \text{otherwise} \end{cases}$$

where $\max = 1$ and $\min = -1$. These two formulae are taken from Rumelhart et al. (1986). After a set number of cycles have passed, weights are adjusted via

$$\Delta w_{ji} = \begin{cases} \alpha(\max - w_{ji})a_i a_j & \text{if } w_{ji} > 0 \\ \alpha(w_{ji} - \min)a_i a_j & \text{otherwise} \end{cases}$$

where the activation values a_i and a_j have been influenced by the preceding recurrent spread of activation, and α is the learning rate (set at 0.1). Via this learning rule, connection weights between nodes will increase to the extent that the nodes have similar activation values, but are constrained not to fall below *min* or above *max*. Following Kruschke (1992), categorization decisions are made by

$$P(c) = \frac{e^{\phi a_c}}{\sum_{k \in K} e^{\phi a_k}},$$

where $P(c)$ is the probability of placing the item in category c , $\phi = 1.0$, K is the set of units representing categories, and a_c is the activation of the node designating category c . Output activations are determined by the same recurrent spread of activation that occurs during training.

(Manuscript received October 6, 1995;
revision accepted for publication February 12, 1996.)