Prototypes, Exemplars, and Object Recognition

Pepper Williams*
University of Massachusetts, Boston

October 16, 1997

Abstract

Four experiments assessed "prototype effects" for six classes of novel 3D objects. As in past categorization experiments, many exemplars of each category were tested, and similarity of test items to category prototypes and to trained exemplars was carefully manipulated. However, as in past object recognition experiments, each category had a unique part structure, making exemplars easily identifiable as one category or another. Old/new recognition performance was strongly related to the similarity between test items and category prototypes, but was unrelated to the similarity between test items and their most similar trained exemplars. A simple neural-network model, trained via a form of Hebbian learning to learn outline contours of the objects, simulated the experimental results surprisingly well. Results are discussed with respect to instance-based and central-tendency models in both the object recognition and categorization literatures.

General Introduction

Visual object recognition—determining the identity of an object from light stimulation on the retinae—is fundamentally both a perceptual and a conceptual problem. That is, successful recognition must involve both perceptual processes that translate raw light information into some more interpretable informational code, and categorization processes that relate this coded information to one object-concept or another. Large literatures have been devoted in cognitive psychology to both halves of this problem. Researchers expressly interested in object recognition have analyzed the perceptual processes involved, while researchers interested in concept formation have analyzed categorization processes. Object recognition researchers regularly point out that the problem is exceedingly difficult because of categorization issues, such as the enormous number of object classes we must know about and the fact that classes overlap each other. Likewise, categorization researchers regularly refer to real-world objects as one of the most important classes of stimuli that must be categorized.

*This research formed part of the author's Dissertation in candidacy for the Ph.D. degree from Yale University. Thanks to Mike Tarr, Jim Anderson, Jim Tanaka, Tim Curran, and Bob Crowder for their comments and support. Correspondence concerning this article should be addressed to: Pepper Williams, Dept. of Psychology, UMass-Boston, 100 Morrissey Blvd., Boston MA 02125; email: *pepper@umbsky.cc.umb.edu*
Version 1.4.

Furthermore, theories of object recognition and theories of categorization have developed along two sets of parallel lines, one line holding that recognition/categorization processes are performed primarily on abstract representations of central tendencies of categories, the other line holding that we process, encode, and retrieve only information about specific encounters with category members. But regardless of whether or not we develop central-tendency representations to categorize objects or other types of stimuli, we must also encode information about individuals in order to later recognize them as such. For example, a building might be recognizable both as "a house" and "Mike's house." The former response could be made on the basis of a central-tendency representation (a.k.a. a *prototype*), but the latter requires a representation of the individual *exemplar* depicted in the picture. Given this consideration, the issue becomes whether exemplar-specific representations are used for both individual-level and category-level recognition decisions, or whether instead prototypes are used for category-level distinctions and exemplars for individual recognition only.

Despite this correspondence in theoretical aims, the object recognition and categorization literatures have developed largely separately. The stimulus set and experiments reported here were designed to incorporate elements of paradigms developed in both literatures, in an attempt to assess the role of exemplars and prototypes in object recognition/categorization processes. In addition, a neural-network simulation model is presented to show how elements of both prototype- and exemplar-based representations can co-exist in a single object recognition/categorization system. Implications of the experimental and simulational results are discussed in relation to theories in both the categorization and object recognition literatures. In the remainder of this introduction, central-tendency and instance based theories in the two literatures are described and the stimulus set and experimental paradigm used here are introduced.

Prototype and Exemplar Categorization Theories

The distinction between prototype- and exemplar-based classification has been most clearly made in the categorization literature. Early experiments in this tradition (e.g., Posner & Keele, 1968) used stimuli such as the dot patterns shown in Figure 1, and required participants to learn to assign training exemplars to two or more categories. The defining result of these experiments was that even though the category prototypes were not seen during training, prototypes were nevertheless categorized

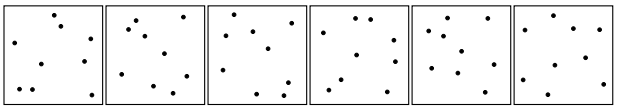

Figure 1: Dot-pattern stimuli like those used by Posner & Keele (1968). The first, fourth, and sixth patterns go in one category; the other three in a second category.

much more accurately and quickly than other untrained exemplars on a later categorization test, and sometimes as accurately as trained exemplars (Posner & Keele, 1970). Posner and Keele, and others after them (Reed, 1972; Homa, Sterling, & Trepel, 1981), took this "prototype effect" to indicate that categorization training results in relatively unstable representations of actual trained exemplars, but stable representations of abstracted central tendencies of the categories.

However, in a landmark paper, Medin and Schaffer (1978) showed that under some circumstances, prototype theories err in their predictions of which test exemplars should be categorized most accurately. Medin and Schaffer, and more recently Nosofsky (1991), Kruschke (1992), and others, proposed that category learners encode trained exemplars only, and that prototype effects can result from the influence of many individually-coded exemplars. This point is elegantly made by Hintzman's (1986) MINERVA2 model. Training exemplars are stored in the model as collections of individual traces, and test exemplars are evaluated by computing the featural match between each test exemplar and all traces. A test exemplar can then be assigned to the category whose encoded members match the test exemplar most closely. MINERVA2 is completely exemplar-based—no prototype abstraction process is ever performed, either explicitly or implicitly, by the model. Nevertheless, MINERVA2 categorizes prototypes better than other untrained exemplars, and when a delay is simulated by randomly changing some of the features in stored exemplars, prototypes are categorized almost as well as trained exemplars.

In fact, prototype- and exemplar-based models become behaviorally indistinguishable (i.e., they predict identical performance patterns) if all trained exemplars are given equal weight in determining test performance (Barsalou, 1990; Estes, 1986). However, most exemplar models (including MINERVA2) give more weight to the most similar, or "nearest-neighbor" trained exemplar. That is, the influence of trained exemplars on test performance is a non-linear function of the similarity between individual training and test items: The greater the similarity, the stronger the influence of the training item (Shepard, 1987). The behavioral consequence of this assumption is that a test stimulus which is highly similar to one trained stimulus and dissimilar to another will be categorized better than a second test stimulus which is moderately similar to both trained stimuli.

Whittlesea (1987) developed a stimulus set and paradigm designed to pit the influence of nearest-neighbor exemplars directly against the influence of prototypes. The stimulus set consisted of five-letter strings in which each exemplar differed by between one and four letters from a category prototype and from each other. For example, NEKAL, a training string in several experiments, differed by two letters from its prototype, NOBAL. One experiment saw participants trained on string such as NEKAL, then tested on strings such as NOKAP, which differs by two letters from both the training string and the prototype, and PEKAL, which differs by only one letter from the training string but by three letters from the prototype. In this and other experiments, Whittlesea found strong evidence for exemplar models, in that similarity to nearest-neighbor training exemplars was a greater determinant of test performance than similarity to prototypes. In the present study, Whittlesea's (1987) stimulus design and paradigm was adapted and extended for the investigation of categorization processes in object recognition.

Viewpoint-Invariant and Viewpoint-Specific Object Recognition Theories

A distinction between models that code specific experiences with stimuli and ones that code abstract representations of stimulus classes is also made, although in different terms, in the object recognition literature. Here, the signature experimental findings are known as *viewpoint invariance* or *viewpoint dependence*. A classic viewpoint-invariant pattern of effects was reported by Corballis, Zbrodoff, Shetzer, and Butler (1978), who showed that letters can be recognized equally quickly in any picture-plane orientation. This type of result may be taken to indicate that abstract object representations are being used to recognize the experimental stimuli. Marr and Nishihara (1978) and later Biederman (1987) proposed more specifically that objects are represented as *structural descriptions*: Viewpoint-invariant descriptions of the simple volumetric "primitives" in an object and how they are related to each other. In Biederman's scheme, a simple house might be coded as something like "short pyramid ABOVE tall cube." Structural descriptions are akin to prototypes in that they are not tied to any one specific encounter with an object. They must be abstracted from one or more encounters, but thereafter one representation serves to recognize the object from any viewpoint.

As in the categorization literature, though, results of other object recognition experiments are inconsistent with theories positing abstract object representations. Tarr and Pinker (1989) showed that recognition of novel letter-like stimuli was dependent on the relationship between test orientations and the orientations in which participants viewed training items. That is, a test stimulus rotated 120° from a trained view took longer to recognize than one rotated 60°. Theories aimed at explaining these viewpoint-dependent patterns of results, like exemplar-based categorization theories, do not posit abstract representations of object classes. Instead, these theories hold that collections of image-based representations, each of which encodes information about an object from a fixed viewpoint, can work together to recognize objects from novel viewpoints (Tarr, 1995; Ullman, 1989). In a final parallel with the categorization literature, structural-description and image-based object recognition theories are often behaviorally indistinguishable, because multiple encoded images of overlearned common objects may lead to the same viewpoint-invariant performance pattern predicted by a structural-description representation.

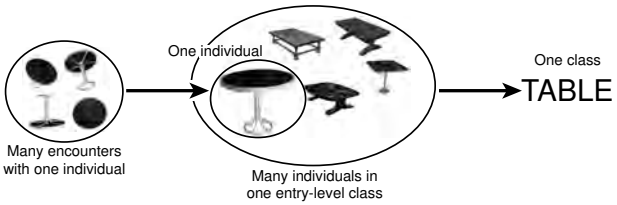

Figure 2: The dual many-to-one mapping problems that must be conquered in object recognition.

The Dual Many-to-One Mapping Problems in Object Recognition

While this discussion highlights important similarities between the prototype/exemplar and structural description/image-based representation distinctions, there are also crucial differences between the approaches taken by categorization and object recognition researchers. Categorization of any set of stimuli (including objects) is difficult because of two intimately related many-to-one mapping problems (Figure 2). First, we may encounter an individual stimulus many times under various circumstances, and the individual stimulus must be considered the same entity in each case. At the same time, however, many individual stimuli must be considered members of the same class. Object recognition researchers have been almost exclusively concerned with the first half of this problem, investigating how images of an object from multiple viewpoints are recognized as the same individual. Categorization researchers have focused instead on the second half of the problem, looking at how different exemplars are recognized as members of the same class.

A second important difference in approaches concerns the "level" of categorization studied. Object recognition research is usually concerned with the *basic* or *entry* level, defined by Rosch and her colleagues as "the level at which categories carry the most information, possess the highest cue validity, and are, thus, the most differentiated from one another" (Rosch, Mervis, Gray, Johnson, and Boyes-Braem, 1976, p. 383[1]). Rosch et al. argued that this level is the first to be accessed when we encounter real-world objects, a view that has proven extremely influential in subsequent object recognition research. Categorization research, on the other hand, is usually concerned with learning of so-called "ill-defined" categories of relatively homogeneous stimuli, such as the dot patterns in Figure 1. Unlike with entry-level object categories, it is not immediately obvious which ill-defined category a given stimulus belongs in.

The present study utilized a mixed approach: As in object recognition paradigms, categories were easily discriminable from each other (Figure 4), but as in categorization paradigms, many exemplars of each category were tested, and exemplars were lawfully related both to each other and to category prototypes (Figure 5). Using this stimulus set, it was possible to test the subtle distinctions between prototype and exemplar categorization

[1] See Jolicoeur, Gluck, and Kosslyn (1984) and Tanaka and Taylor (1991) for important qualifications on the idea that a single basic level exists for all objects and all observers, and for why the term entry-level is therefore to be preferred.

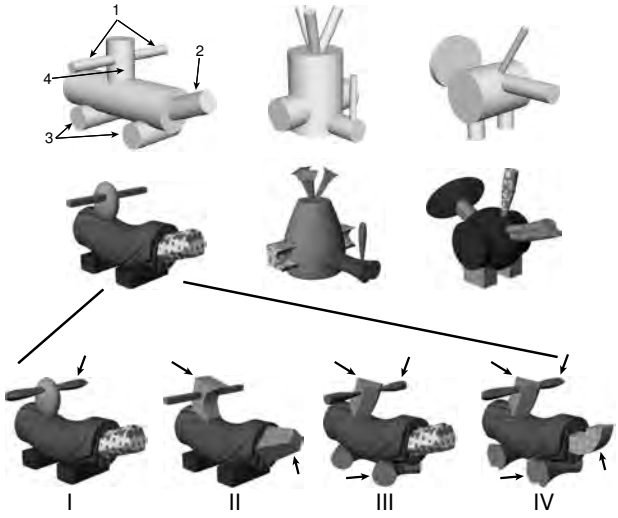

Figure 3: Part structures (top row) and prototypes (middle row) for three Fribble species, and four more exemplars (bottom row) of one species.

theories in the context of entry-level object recognition. Results indicate that a mixture of prototype-based and exemplar-based processes are necessary to account for the recognition of entry-level object categories.

Stimulus Set Design

Figure 3 shows the basic design for the set of objects (dubbed *Fribbles*) constructed for the experiments (the objects were designed and rendered using StudioPro software [Strata Inc., St. George, UT] on a Macintosh computer). Six Fribble *species* were used, where each species was defined by a unique part structure that consisted of a large *main body* with four *appendage parts* attached. The top row of Figure 3 shows abstract part structures for three species; note that some appendages, such as the parts numbered 1 and 3 in the top-left picture in the Figure, were repeated symmetrically two to four times in an object. Between-species variation was qualitatively similar to the variation between entry-level object categories (e.g., airplanes and cars).

Individual Fribble exemplars were created by substituting richly colored and textured three-dimensional volumes for the gray cylinders used in designing the part structures. Within a species, the following object attributes were held constant across all exemplars: the exact shape of the main body, the approximate location and interrelationships between the appendage parts, and the approximate colors and textures of the appendage parts. The main aspect that varied from exemplar to exemplar in a species was the exact shapes of the appendage parts (bottom row of Figure 3). Each appendage "socket" could take one of three different three-dimensional volumes. For example, the first three exemplars in the bottom row of the Figure show the three possible volumes that could be used as the appendage part numbered 4. All told, there were 81 potential exemplars from each species.[2]

[2] An elaborate counterbalancing scheme was devised to assure that no one three-dimensional volume was perfectly diag-

One exemplar of each species was arbitrarily designated the prototype for the species; Figure 4 shows the six prototypes. A naming convention (derived from Whittlesea, Brooks, & Westcott, 1994) for the Fribbles was adopted in which an initial Roman numeral always designated the number of parts differing between the exemplar and the species prototype. The bottom row of exemplars in Figure 3 shows I, II, III, and IV exemplars of the species whose prototype is shown on the left of the middle row; arrows indicate the parts changed from the prototype for each exemplar.

Besides the prototypes, six other sets of exemplars were constructed for each Fribble species. Eight exemplars each differing by two parts from the species prototypes served as training items in Experiments 2-4. This set was designated pII-e0, where the second, Arabic numeral indicates how many parts differ from the nearest (most similar) trained exemplar (the Arabic numeral is zero for this set since these will be the trained exemplars themselves). This set "surrounds" the category prototypes, in that across the eight pII-e0 exemplars, the modal appendage in each socket corresponds to the prototype (more precisely, the prototype shape for each appendage is present in four exemplars and each of the other two shapes is present in two exemplars). Figure 5B shows this relationship schematically, and Figure 5A shows the actual objects in the set for one Fribble Species.

Other item sets vary in systematic ways from both the species prototypes and the nearest trained exemplars. Items in the pI-e1 and pIII-e1 sets differ by one part from the nearest trained exemplar and by one and three parts, respectively, from the species prototype. Similarly, pII-e2 and pIV-e2 items differ by two parts from the nearest trained exemplar and by two and four parts from the prototype. Again, Figure 5B shows the relationship between sets schematically, and Figure 5A shows the complete sets for one species. Prototypes can also be labeled p0-e2 items, since they differ by zero parts from the prototype and by two parts from the nearest trained exemplars.

Note that the number of parts differing from the nearest trained exemplar is not the same as the *average* number of parts different from *all* pII-e0 exemplars. For example, each pII-e2 item differs by two parts from five pII-e0 items, by three parts from two other pII-e0 items, and by four parts from the remaining pII-e0 item. On average, then, each pII-e2 item differs by $((2 \times 5) + (3 \times 2) + (4 \times 1))/8 = 2.5$ parts from all pII-e0 items. The same computations for other item sets give average distances of 2.0 parts for p0-e2 items, 2.25 parts for pI-e1 items, 2.5 parts for pII-e0 items, 2.75 parts for pIII-e1 items, and 3.0 parts for pIV-e2 items.

Overview of Experimental Paradigm

The experiments reported here used these objects in a paradigm taken directly from the categorization literature (Franks & Bransford, 1971). Participants first studied one item set from each Fribble species (training items), then were tested on this and other item sets that were not studied ("transfer items;" Posner and Keele, 1968). The training task varied from experiment to experiment, but usually involved learning to categorize Fribbles at the species level; for example, learning that all the items in Figure 5 are members of the SOGI species. The test task employed was an old/new recognition test, in which participants saw objects one at a time and decided whether or not each one was seen during training. Participants also rated their confidence on each decision, and the two judgments (old/new and confidence) were combined to produce a confidence score that ranged from 1 (very confident new) to 6 (very confident old).

Instructions emphasized that a test item had to be *exactly* the same as a training item to be called OLD, so analyses focused on differences between confidence scores for various types of transfer items of each species. In Experiments 2-4, when participants were trained on pII-e0 items, the classical prototype effect (Posner & Keele, 1968) would be indicated by higher confidence scores for prototypes than for other untrained transfer items.

This and other potential predictions will be fleshed out below. In Experiment 1, however, a simpler prediction was tested: It was possible that the exemplars within each species were so similar to each other that participants would completly ignore intra-species variation and give equivalent confidence scores to all items. Participants in this experiment were trained to categorize the six species of Fribbles, then received a recognition test on prototypes, pI-e1, pII-e2, and pIII-e1 items. Unlike in Experiments 2-4, participants in Experiment 1 were trained on species prototypes, instead of on pII-e0 items (the Arabic numerals in item set names are not accurate in the context of Experiment 1, since they refer to distance from pII-e0 items; items will therefore be referred to solely by their Roman numerals in this experiment). This experiment was thus designed so that almost any categorization or object recognition model would predict the same pattern of effects—confidence scores should vary with distance from the species prototype, since this item is not only the prototype but also the sole trained exemplar. Such a pattern would demonstrate that meaningful performance differences are obtainable from different exemplars and also provide a baseline for expected performance in Experiments 2-4.

Experiment 1

Method

Participants. Thirty-two Yale or Brown University undergraduates participated in exchange for course credit or payment.

Materials. The design and method for constructing the Fribble stimuli are described in the General Introduction. Four-letter pseudowords (starting with the letters S, D, F, J, K, L, for convenience in performing keyboard-response naming tasks) were assigned to each species; name-species assignments are shown in Figure 3. All experiments were run on Macintosh computers. Fribbles were always displayed one at a time, and appeared within a 400-by-340-pixel area in the middle of the computer screen. Pixel

nostic for any species. For example, the side-attached parts in the object in the center of the middle row of Figure 3 share the same shape with part 4 in the object second from the left in the bottom row. Color/texture combinations were also repeated in more than one species. Furthermore, two species shared each main-body shape (that is, there were three unique main-body shapes for the 6 species).

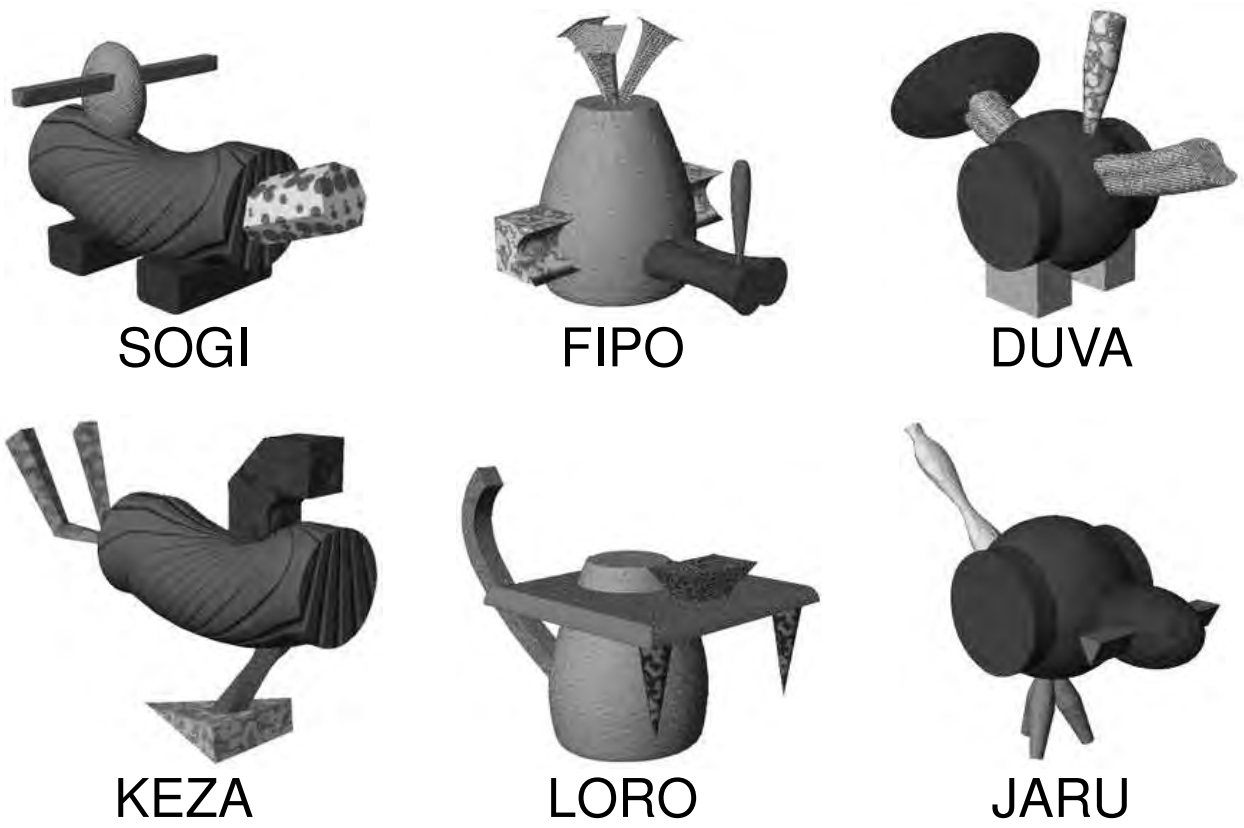

Figure 4: Prototype exemplars for the six Fribble species. Also shown are the nonsense-word names assigned to the species in Experiments 1-3. Actual stimuli were brightly colored and texture differences were more pronounced.

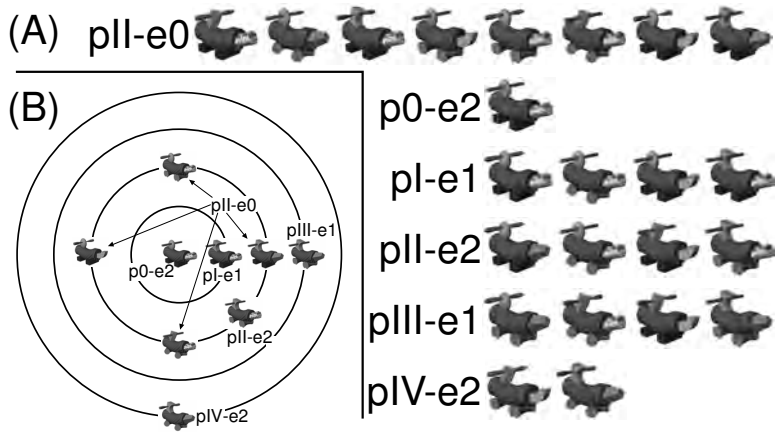

Figure 5: Object sets used in the experiments. A) The twenty-three exemplars used for one Fribble species. B) Schematic representation of the relationships between sets.

sizes were approximately 28 or 22 pixels/cm, and participants sat about 60 cm from the screen, resulting in maximum visual angles of approximately 13.2° by 11.3° or 17.0° by 14.6°.

Procedure. The experimental session, which lasted between 25 minutes and an hour in this and all subsequent experiments, consisted of a training phase followed by a test phase. The goal of the training phase was to learn names for the six Fribble species. Only the prototype of each species was shown in training. Procedures for all training and test tasks, including intertrial intervals and fixation durations, are summarized in Table 1.

In the initial *Train-View* task, participants were told to simply "look at each object and try to remember its category name." Each of the six training Fribbles was shown one at a time with its name printed underneath it, and this series of six trials was repeated three times, each time in a different random order. In the second, *Train-Press* task, participants saw the same display as in the first task, but were now required on each trial to press the key corresponding to the first letter of the object's name. Again, the series of six objects was shown three times, in different random orders. In trials of the third, *Train-Feedback* task, participants first saw an object without its name, and attempted to name the object (i.e., pressed the key that they thought corresponded to the first letter in the object's name). If the response was incorrect, the object was shown again, this time with its name; if the response was correct, the next trial began immediately. Again, the series of six objects was shown three times, in different random orders.

In the fourth task, *Train-Test*, participants saw an object without its name, attempted to name the object, and heard a "beep" if they were incorrect, or heard nothing if correct. Once again, the series of six objects was shown three times, in different random orders. Following the initial Train-Test task, participants performed a block of six Train-Feedback trials, then another block of 18 Train-Test trials. If a participant got 17 or 18 of the Train-Test trials correct, they were allowed to proceed to the test phase; otherwise, they were required to repeat the six-trial Train-Feedback and 18-trial Train-Test tasks until this criterion was met, or until they had completed 6 Train-Test tasks.

In the test phase, which immediately followed the completion of training, participants performed the Recognition test, in which they decided whether or not test objects had been seen during the training phase, then rated their confidence in the decision (a 500-ms blank screen intervened between the old/new and confidence judgments). Instructions emphasized that the object had to be exactly the same as a training item to be called OLD, because many objects would look similar but be slightly different from the training objects. Old/new responses were made by pressing keyboard keys B for OLD or N for NEW; confidence ratings were made by pressing keyboard keys 1 ("I think it MIGHT have been/not have been one that I saw before"), 2 ("I think it PROBABLY was/wasn't one that I saw before"), or 3 ("I'm CERTAIN it was/wasn't one that I saw before"). Recognition test stimuli included the prototypes and one object from each of the I, II, and III sets, for each of the six Fribble species (II items were drawn from the pII-e2, not the pII-e0, set). The specific item used from each set was counterbalanced across participants. The correct response was OLD for the prototypes and NEW for the other test items; thus the ratio of OLD:NEW responses was 1:3 for the 24 test trials.

Design and statistical analyses. The independent variable for the Recognition task was distance from prototype (0, I, II, or III parts), manipulated within-participants. The dependent measure reported is a confidence score, calculated by combining the old/new and confidence rating responses so that 1 = high confidence new, 2 = medium confidence new, 3 = low confidence new, 4 = low confidence old, 5 = medium confidence old, and 6 = high confidence old. Separate analyses on raw proportions of OLD responses yielded almost identical patterns of effects, but were slightly more variable. Because Recognition accuracy levels were often close to chance (especially in later experiments), response times were difficult to interpret, and are not reported. All reported inferential statistics were significant at the .05 level, unless noted.

Results and Discussion

Training. Participants were required to name 17/18 Fribbles correctly on a Train-Test in order to go on to the test phase. All participants met this requirement; in fact, 24 of the 32 participants (75%) achieved this level of performance on the very first Train-Test, after having seen each species prototype nine times. The percentages of participants reaching criterion by the second, third, fourth, and fifth Train-Tests were 84%, 88%, 94%, and 100%.[3] For the first four Train-Feedback and Train-Test tasks, which all participants completed, mean accuracy rates were .84, .92, .91, and .95. The extremely good performance on even the first Train-Feedback test attests to the high discriminability between Fribble species.

Recognition. Mean confidence scores for the Recognition test are shown in Figure 6. A significant linear contrast test, $F(1, 93) = 321$, $MS_e = 0.47$, indicated that confidence scores varied strongly according to distance from category prototypes. The slope of the best-fitting regression line was -0.98 confidence units/part (that is, changing one part from the prototype led to a 0.98-unit decrease in confidence). Clearly, Fribble exemplars are different enough from each other to produce stable and interpretable patterns of effects on the old/new Recognition test.

Experiment 2

Recognition confidence scores in Experiment 1 varied regularly with the distance (dissimilarity) between novel transfer items and category prototypes. This finding does not, however, indicate a "prototype effect," because participants were trained on the prototypes themselves. In Experiment 2, participants were trained on the eight pII-e0 exemplars of each species that surround the species

[3] Several of the participants who took four to five tests to reach criterion were hampered by the fact that because of an experimenter error, the sound was turned off on the computers they were using; thus they received no feedback on the Train-Tests.

Table 1: Summary of procedures in training and test tasks.

Task Label	ITI/Fix/SD[a]	Task Description	Feedback
Train-View	500/0/5000	See object with name; just look.	None.
Train-Press	1000/500/free[b]	See object with name; press name.	None.
Train-Feedback	1000/500/free	See object; press name.	See object w/ name for 3000 ms if incorrect.
Train-Test	1000/500/free	See object; press name.	Beep if incorrect.
Recognition	1500/500/free	See object; decide if seen during training; rate confidence of decision (1-3 scale).	None.

Note. [a]ITI = intertrial interval; Fix = duration of fixation cross; SD = stimulus duration. [b]free = free view (object stays on screen until a response is made).

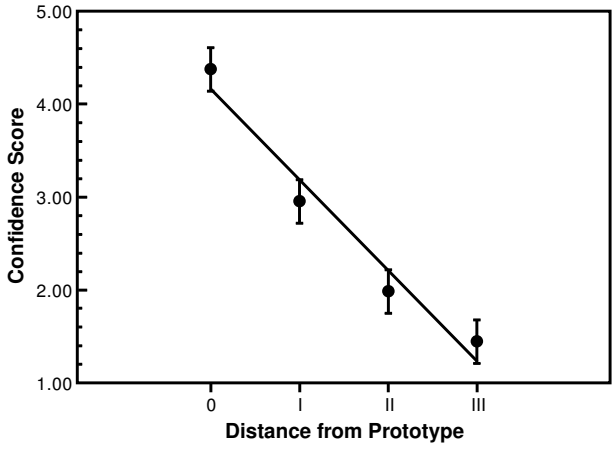

Figure 6: Mean Recognition confidence scores in Experiment 1. The line shows the best-fitting relationship between confidence scores and distance from prototype; error bars show within-subject confidence intervals (Loftus & Masson, 1994).

prototypes (Figure 5B). The training tasks, as in Experiment 1, taught participants to categorize Fribble exemplars into species. If, during training, participants extract the species prototypes from these training exemplars, then the same pattern of effects would be predicted for this experiment as for Experiment 1: Confidence scores should vary along with distance from the prototype, given by the Roman numeral in item-set names. This pattern is diagrammed in Figure 7A.

Note that there are two potentially independent prototype effects in this pattern. First, the prototype may be responded to with greater strength than trained exemplars, and second, response strength may diminish as an item's distance from the prototype increases. For ease of discussion, the former will be referred to as the *prototype-advantage effect*, and the latter as the *prototype-distance effect*. The strength of the prototype-distance effect is indicated by the slope relating distance from the prototype with confidence scores for untrained transfer items; the strength of the prototype-advantage effect is simply given by the pairwise comparison of confidence scores for prototype and pII-e0 items. Previous studies using categorization test tasks (Posner & Keele, 1968) have generally found strong prototype-distance effects, but weaker prototype-advantage effects. Experiments employing old/new recognition tests (Franks & Bransford, 1971) generally show stronger prototype-advantage effects (but see Homa, Goldhardt, Burruel-Homa, & Smith, 1993). A prototype-advantage effect in the present paradigm would also be reminiscent of results from the "false memory" paradigm recently revitalized by Roediger and McDermott (1995), in which study of word lists containing strong associates (e.g. THREAD, PIN, EYE, SEWING) of an unstudied critical lure (NEEDLE) leads to high levels of false recall and/or false recognition of the critical lure.

As noted above, the pattern shown in Figure 7A would also be predicted by an exemplar-based categorization model that weighted all training exemplars equally, but most exemplar-based models (Hintzman, 1986; Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1991) assume non-linear weighting of exemplars, such that trained items highly similar to a test item will have larger effects on test performance than less-similar trained items. This type of model might predict the pattern of effects shown in Fig-

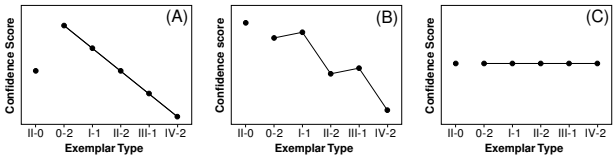

Figure 7: Three potential patterns of effects in Experiments 2-4: A) performance is related solely to distance from species prototypes; B) performance is related both to distance from prototypes and distance from nearest trained exemplars; C) performance is unrelated to distance from either prototypes or nearest trained exemplars.

ure 7B, in which performance is predicted both by distance from the species prototype and by distance from the nearest (and thus most strongly weighted) training exemplar. This pattern would be consistent with the data reported by Whittlesea (1987) in the experiments discussed above that formed the basis for the present stimulus-set design.

A third possibility is that participants will extract and remember only the abstract part-structures of Fribble species (top row of Figure 3) during training, leading to equivalent confidence scores for all types of transfer items. This pattern is diagrammed in Figure 7C. While participants in Experiment 1 did not base test performance on abstract part-structures (since confidence scores did vary by item type), such representations would make more sense in the context of Experiment 2, where eight different exemplars have to be associated with each species: An abstract representation that does not specify exactly what three-dimensional shape goes in each part socket could actually be more efficient than a representation of the category prototype for categorizing the various exemplars in this experiment.

Method

Participants. Twenty-four undergraduates from Brown University or Oberlin College served as participants in exchange for payment.

Materials. Presentation of Fribbles was the same as reported in Experiment 1. Training stimuli were the eight pII-e0 exemplars of each of the six Fribble species. These exemplars were designed so that if test performance in the main experiments relies on central-tendency representations abstracted from trained items, or on the average distance of a test item from all trained items, prototypes should show the greatest response strengths. Another way to state this assumption is in terms of a "similarity space," where points representing objects are plotted in such a way that the physical distance between any two points is inversely proportional to the perceived similarity between the objects. For each Fribble species, the prototype should lie at the center of such a space, and the pII-e0 items should encircle it; an idealized similarity space is depicted in Figure 5B.

Although the pII-e0 set was constructed so that the prototype was, objectively, the object containing the modal features of the pII-e0 items, a pilot experiment was run to verify that subjective impressions of human participants would agree with this objective stimulus-space

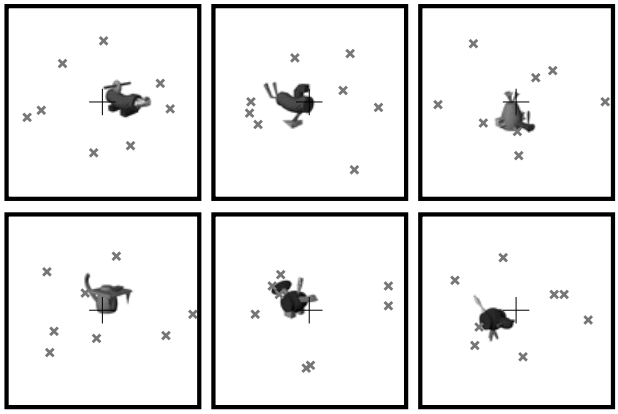

Figure 8: Multi-dimensional scaling solutions for similarity ratings of exemplars from six Fribble species. The picture in each plot represents the prototype of the species; pII-e0 items are represented as x's.

structure. Twelve participants made pair-wise similarity judgments for the pII-e0 items and prototype of each species, and a multi-dimensional scaling (MDS) solution was constructed, on the basis of these ratings, for each species. As shown in Figure 8, the prototype was the item nearest to the center of each solution. In addition, participants also made typicality judgments on the nine items from each species, and the prototype was rated more typical than any of the pII-e0 items for all six Fribble species.

Procedure. Unlike in Experiment 1, where participants were trained to a criterion, participants in this and subsequent experiments received a fixed amount of training, regardless of how proficient they became. Aside from this change, training was similar to Experiment 1, in that participants learned to categorize Fribbles into species via the Train-View, Train-Press, Train-Feedback, and Train-Test tasks (Table 1. Training protocols (sequences of training tasks) are given in Table 2. Twelve participants performed the Long training protocol, while the other 12 performed the Short protocol.

Immediately following completion of the training protocol, participants received a Recognition test (Table 1), the procedure for which was the same as in Experiment 1. Test stimuli included, for each Fribble species, the eight pII-e0, one prototype, four pI-e1, four pII-e2, four pIII-e1, and two pIV-e2 exemplars (all the exemplars shown in Figure 5A). The correct answer was OLD for pII-e0 items (35% of test trials) and NEW for all others (65% of trials). Participants were informed of this ratio of old-new trials.

Design and analyses. In this and subsequent experiments, exemplar type (pII-e0, prototype, pI-e1, pII-e2, pIII-e1, or pIV-e2), manipulated within-participants, was the primary independent measure. Training protocol (Long or Short) was also an independent measure in Experiment 2, manipulated between-participants. Confidence scores were obtained as in Experiment 1 and used as the dependent measure. An omnibus analysis of variance (ANOVA) was run to determine if performance varied significantly from the abstract structure pattern (Figure 7C). To assess the fit of the prototype vs. nearest-

Table 2: Training protocols and training performance in Experiment 2.

Task	Ex[a]	N[b]	Acc[c]
Long Protocol			
Train-View	1-4	24	
Train-Press	1-4	24	
Train-View	5-8	24	
Train-Press	5-8	24	
Train-Feed.	1-4	24	.90
Train-Test	1-8	48	.96
Train-Feed.	5-8	24	.97
Train-Test	1-8	48	.96
Total Trials		240	
Short Protocol			
Train-View	1-4	24	
Train-Press	1-4	24	
Train-Press	5-8	24	
Train-Feed.	1-4	24	.90
Train-Test	1-8	48	.97
Total Trials		144	

Note. All training in this experiment was at the species level. [a]Ex = exemplars included in task; [b]N = total number of trials in task; [c]Acc = accuracy rate for Train-Feedback and Train-Test tasks.

neighbor patterns (Figure 7A and B), a least-squares regression analysis was run on confidence scores for unstudied item types, with distance from prototype and distance from nearest trained exemplar as predictors. Both the prototype and nearest-neighbor patterns predict a significant regression coefficient for distance from prototype (this coefficient also gives the slope of the prototype-distance effect), but the nearest-neighbor pattern additionally predicts a significant coefficient for distance from nearest trained exemplar. The strength of the prototype-advantage effect was measured by a pairwise comparison, including t and sign tests, between pII-e0 items and prototypes. All ANOVAs and regression analyses were performed on participant means.

In the Short training protocol, one half of the training items (exemplars 1-4 of the pII-e0 set) were seen four times each, while the other half (exemplars 5-8) were seen only twice. To better compare the two protocols, only exemplars 1-4 of the pII-e0 set were included in the main analyses, and to facilitate comparisons across experiments, exemplars 5-8 were also excluded from analysis in subsequent experiments (however, results from both sets of training exemplars are considered below in connection with the neural-network model).

Results and Discussion

Training. Performance on Train-Feedback and Train-Test tasks during training was again very good, as shown in Table 2. Performance in the present experiment is not completely comparable to that in Experiment 1, because training in this and subsequent experiments included eight, rather than only one, training exemplars per species. Nevertheless, it is obvious that participants once again quickly became proficient at categorizing the Fribble species.

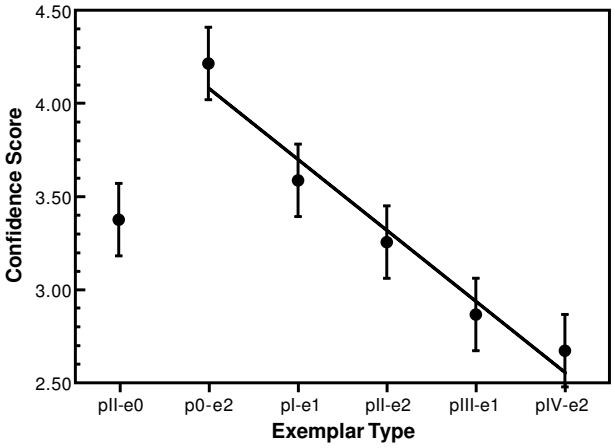

Figure 9: Mean Recognition confidence scores in Experiment 2. The line shows the best-fitting relationship between confidence scores and distance from prototype; error bars show within-subject confidence intervals (Loftus & Masson, 1994).

Recognition. Confidence scores for the Recognition test are shown in Figure 9. An initial ANOVA including the training protocol factor revealed neither a significant main effect of protocol nor a significant interaction between protocol and exemplar type (both $F < 1$), so all participants' results were combined in the graph and in subsequent analyses.

Participants in Experiment 2 clearly remembered something about the specific objects they saw during training. If they had simply developed abstract structural-description representations of the Fribble species, they should have shown no differences between the various test sets (Figure 7C), a hypothesis soundly refuted by the omnibus ANOVA, $F(5, 115) = 30.7$, $MS_e = .236$. Figure 9 also shows no evidence whatsoever for the nearest-neighbor exemplar pattern of effects (Figure 7B): Confidence scores varied with distance from species prototypes, but were not affected at all by distance between test exemplars and their nearest trained exemplars. Supporting this conclusion, the regression analysis showed a significant coefficient for distance from prototype, $t(117) = 9.33$, but not for distance from nearest trained exemplar, $t(117) = 1.32$.

Furthermore, participants were more confident that they had seen the prototypes of Fribble species than that they had seen the actual items on which they were trained (the difference in confidence scores was .84, $t(23) = 4.80$, and 88% of participants were more confident on prototypes than pII-e0 items, $z = 3.67$). These results indicate that the data is well-explained by either of two types of models: A prototype model that abstracts the modal feature values for each Fribble species, rather than abstract structural-descriptions (Reed, 1972), or an exemplar model that considers all trained items more or less equally when making decisions, rather than giving special treatment to the nearest trained exemplar.

The regression analysis indicated that each part of a test exemplar that differed from the prototype led to a

.381 decrement in confidence scores. This slope was substantially shallower than the one found in the Recognition task of Experiment 1 (-.976). A possible danger in interpreting these results is that because of the many (138) trials in the Recognition test, participants may have shown the steep slope because of learning that went on during the test itself. That is, they may have started out showing a much shallower prototype-distance effect, or even a nearest-neighbor pattern, but the latter effect could have been washed out in later trials if prototypes were developed during the test itself. However, an additional analysis revealed that the slope was actually steeper in the first half (-.432 confidence units/part) than in the second half (-.359) of trials, effectively ruling out this possibility.

The results of this experiment are somewhat surprising given the fact that Whittlesea (1987), using a similar paradigm and stimulus-set design, found strong evidence for nearest-neighbor matching processes. However, the training in Experiment 2 focused participants exclusively on the *inter*-species differences between items, forcing them to ignore all *intra*-species differences. While this procedure is typical of many categorization studies, it is unlike what we usually experience when learning about real-world objects, for we must encode information both about categories and specific individuals. Moreover, most of Whittlesea's (1987) experiments involved a training task (rote copying of letter strings) that required no inter-category comparisons whatsoever.

In Experiment 3, participants were led to more closely examine intra-category differences by requiring them to learn *owner names* for some of the training exemplars. That is, they had to learn not only that some objects were SOGIs, but also that one of the SOGIs was owned by Carlos, another by Vera, and so on. If the lack of nearest-neighbor effects in Experiment 2 was due to a lack of attention to details of the training exemplars, this procedural change should cause a shift towards the nearest-neighbor pattern of performance.

Experiment 3

Method

Fifteen Brown University or Oberlin College undergraduates served as participants in exchange for course credit or payment. Materials, as well as procedures and design for the test phase, were identical to Experiment 2. During training, participants attempted to learn both the species names for all eight pII-e0 exemplars of each species, and owner names for four of the eight exemplars. The training protocol is given in Table 3. The first 18 tasks[4] included objects from only one Fribble species each. In these blocks, participants were informed of the species-level name for the species, and tried to learn owner names for each of four exemplars of the species. For the SOGI species, for example, participants were told that one object was "Carlos' SOGI," one was "Vera's SOGI," one was "Nancy's SOGI," and one was "Marvin's SOGI." The owner names were repeated across species; thus participants learned that Nancy owned one SOGI, one DUVA, one FIPO, etc. The order in which species were learned was randomized for every participant. The nineteenth and twentieth tasks tested owner names for all six categories together. Next, participants were further trained on species-level names in a similar fashion as in Experiment 2 (Table 3), and were finally given one last test on owner names.

[4] Actually, these 18 tasks were preceded by three in which participants learned owner names for a practice species, that was not included in the test phase.

Table 3: Training protocol and training performance in Experiment 3.

Task	Level	Ex[a]	N[b]	Acc[c]
Train-View	Owner	1-4	4	
Train-Press	Owner	1-4	4	
Train-Feedback	Owner	1-4	12	
Repeat the previous three tasks six times, once for each species.				
Train-Feedback	Owner	1-4	24	.67
Train-Test	Owner	1-4	24	.72
Train-Press	Species	5-8	24	
Train-Feedback	Species	5-8	24	.87
Train-Test	Species	1-8	48	.93
Train-Feedback	Species	5-8	24	.94
Train-Test	Species	1-8	48	.95
Train-Test	Owner	1-4	24	.68
Total Trials			360	

Note. [a]Ex = exemplars included in task; [b]N = total number of trials in task; [c]Acc = accuracy rate for Train-Feedback and Train-Test tasks.

Results and Discussion

Training. Performance on Train-Feedback and Train-Test tasks during training is shown in Table 3. Owner-level names were much more difficult than species-level names to learn: The final Train-Test task followed seven trials on each exemplar, and participants still managed only 68% correct. This is well above the chance level of 25%, but significantly below the very high (95%) level of performance participants obtained in the immediately previous species-level Train-Test ($t(14) = 5.42$; all 15 participants were better on the species than on the owner test). Participants were also significantly slower on owner-level trials (mean 1920 ms) than on species-level trials (861 ms, $t(14) = 8.71$; again, all 15 participants showed the same effect).

Recognition. Confidence scores for the Recognition test are shown in Figure 10. As in Experiment 2, performance varied significantly by exemplar type, $F(5, 70) = 10.8$, $MS_e = .230$, and the regression coefficient was significant for distance from prototype, $t(72) = 4.31$, but not for distance from nearest trained exemplar, $t < 1$. However, the confidence score/prototype-distance slope in Experiment 3 (-.264 confidence units/part) decreased relative to Experiment 2. Note that this shallower slope was not due to a decrease in memory for trained objects—compared to Experiment 2, confidence scores in Experiment 3 went up more for pII-e0 items (an increase of .59) than for other items combined (an increase of .39), indicating greater discriminability of studied and unstudied items in the present experiment.

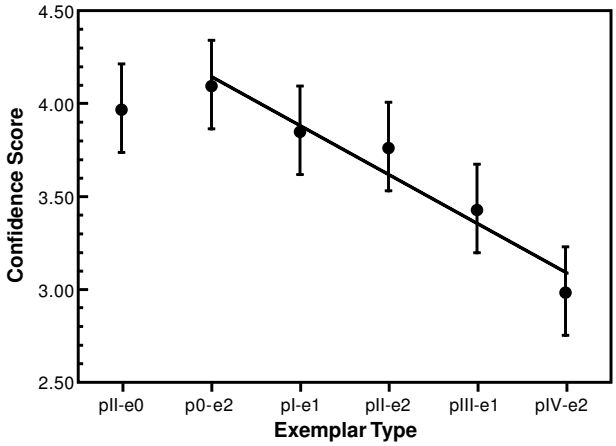

Figure 10: Mean Recognition confidence scores in Experiment 3. The line shows the best-fitting relationship between confidence scores and distance from prototype; error bars show within-subject confidence intervals (Loftus & Masson, 1994).

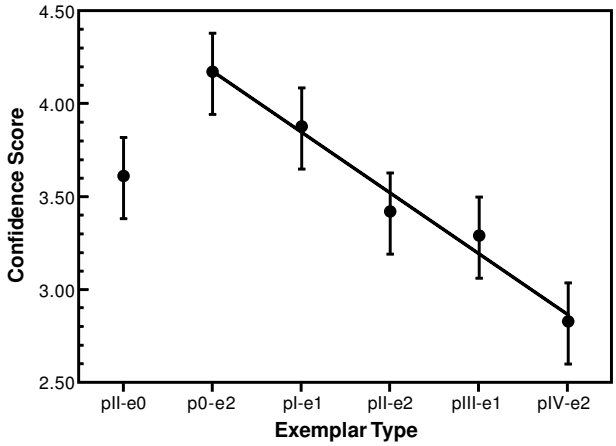

Figure 11: Mean Recognition confidence scores in Experiment 4. The line shows the best-fitting relationship between confidence scores and distance from prototype; error bars show within-subject confidence intervals (Loftus & Masson, 1994).

Another important change in the present experiment was that the prototype/pII-e0 difference, while in the same direction as in the previous experiment, was much smaller (.13 confidence units), and did not reach significance by either a t- ($t < 1$) or sign test (60% of participants were more confident on prototypes than pII-e0 items, $z < 1$). A between-experiments analysis indicated that the prototype/pII-e0 difference was significantly larger in Experiment 2 than in Experiment 3, $F(1, 37) = 5.75$, $MS_e = 0.819$.

Thus the addition of owner-level training did allow participants to retain enough specific information about the exact exemplars they were trained on to be able to recognize them with almost as much confidence as prototypes. Nevertheless, the nearest-neighbor pattern of performance again failed to materialize in the untrained item sets. One possible interpretation of Experiments 2 and 3 is that any kind of study task that requires participants to learn to classify stimuli, regardless of the level of classification, leads participants to encode information about individuals relative to other Fribbles. This type of encoding could in turn lead participants to take all trained items into account equally when performing the Recognition test. If so, then perhaps a study phase that includes no classification training at all will produce the nearest-neighbor effects found by Whittlesea (1987).

This was the motivation for Experiment 4, in which participants performed four incidental-learning study tasks (Table 4), then received the same recognition test as in the previous three experiments. The study tasks were the standard kinds of rating tasks used in countless memory experiments (e.g., Schacter, Cooper, & Delaney, 1990).

Experiment 4

Method

Fifteen Brown University undergraduates participated in exchange for course credit or payment. Materials, as well as procedures and design for the test phase, were identical to Experiment 2. During training, participants learned neither species names nor owner names. Instead, they performed a series of rating tasks on the same pII-e0 items learned in Experiments 2 and 3. The training protocol, along with descriptions of the rating tasks, are given in Table 4. Note that although the tasks were completely different, the structure of the protocol (i.e., the order in which exemplars were introduced and the number of exposures to each exemplar) was exactly the same as the Short protocol shown in Table 2 and used in Experiment 2.

Results and Discussion

Recognition. Confidence scores for the Recognition test are shown in Figure 11. Once again, performance varied significantly by exemplar type, $F(5, 70) = 19.0$, $MS_e = .171$, the regression coefficient for distance from prototype was significant, $t(72) = 5.35$, and the coefficient for distance from nearest trained exemplar was not significant, $t < 1$. The slope in Experiment 4 (-.327) jumped back up to a level in-between Experiments 2 and 3, and the magnitude of the prototype/pII-e0 difference was also in-between the previous two experiments (.56 confidence units, $t(14) = 3.03$; 67% of participants were more confident on prototypes than pII-e0 items, $z = 1.94$).

Participants in Experiment 4 were not asked to categorize the Fribbles in any way. Therefore, there would have been no motivation for them to consciously attempt to abstract species prototypes during training. Nevertheless, participants demonstrated even stronger prototype effects (i.e., a steeper slope and larger prototype/pII-e0 difference) than in Experiment 3, in which species categorization was required. This result indicates either that prototype abstraction is an automatic, ubiquitous process, or that some other mechanism besides prototype abstraction lies at the root of the prototype effects found in Experiments 2-4.

Table 4: Training tasks and protocol in Experiment 4.

Task Name	Ex[a]	N[b]	Description
Part-Count	1-4	24	Count the number of parts in the object.
Pleasantness	1-4	24	Rate how "aesthetically pleasing" each object is (1 = unpleasant, 9 = pleasant).
Pleasantness	5-8	24	Same as above.
Organicness	1-4	24	Rate how "organic" or "man-made" each object looks (1 = organic; 9 = man-made).
Distinctiveness	1-8	48	Rate how "distinctive" each object looks (1 = neutral; 7 = distinctive).
Total Trials		144	

Note. [a]Ex = exemplars included in task; [b]N = total number of trials in task.

MINERVA2 Simulations

In fact, Hintzman's (1986) MINERVA2 model provides such an alternative mechanism. As noted previously, this model produces prototype effects even though memory traces of trained exemplars are stored separately in the model's memory and no prototype abstraction process is posited. Here, MINERVA2 simulations of Experiments 2 and 3 were attempted using Hintzman's (1988) simulations of other recognition experiments as a guide. The *echo intensity* of a test probe, analagous to the confidence score given to a test item in the present experiments, was determined by summing the activation of all memory traces in response to the probe (probes and traces were instantiated as collections of features that can take on values of +1 and -1[5]). Activation of a trace was in turn based on the similarity of the trace with the probe. Specifically, similarity between a probe P and a trace T_i, each of which are made up of N features, was computed by the formula

$$\mathrm{S}_i = \sum_{j=1}^{N} P_j T_{i,j} / N,$$

where P_j was the value of the jth feature of the probe and $T_{i,j}$ was the value of the jth feature of the trace. Activation was then given by

$$\mathrm{A}_i = \mathrm{S}_i{}^3.$$

Defining activation (and thus echo intensity) as a positively accelerated function of similarity not only allows MINERVA2 to distinguish learned items from similar distractors, but also conforms to the doctrine held by most exemplar-based models that "the influence of an exemplar drops off rapidly with decreasing similarity to the probe" (Hintzman, 1986, p. 424).

In an initial simulation using the MINERVA2 model, exemplars were coded as collections of 12 features, where each feature-triplet represented an appendage-part socket (simulations coded and tested only one species; it was assumed that inter-species variation was great enough that feature-vectors representing exemplars of different species would be roughly orthogonal to each other). The prototype shape for an appendage was coded as {+1, -1, -1}, and the alternate shapes as {-1, +1, -1} and {-1, -1, +1}. Traces were constructed for the eight pII-e0 training items, then probes representing items of all exemplar types were compared to the traces via the above equations to produce echo intensities. Results of this simulation are shown in Figure 12A, which strongly resembles the nearest-neighbor pattern of effects diagrammed in Figure 7B.

This pattern decribes fairly well the results of Whittlesea's (1987) experiments, but does not agree with the present findings, in which no nearest-neighbor effects were found. However, representing Fribbles solely in terms of the features on which they differ from exemplar to exemplar is unrealistic: Within a species, exemplars have more common aspects than distinguishing aspects. In a second MINERVA2 simulation, this intra-species fraternity was represented by adding 48 features that were each identical from exemplar to exemplar. This simulation produced a pattern of effects (shown in Figure 12B) very similar to that found in Experiment 2.

The source of this pattern shift can be seen by examining the individual similarity scores, and more importantly, the activation values, of each of the eight pII-e0 exemplars with prototype and pII-e0 test probes (relevant activation values are given in the insets of Figure 12). In the first simulation (with no common features), the activation contributed by a pII-e0 trace to its identical probe was roughly 25 times the activation contributed by this trace to the prototype, far outweighing the smaller activation advantage of the prototype over the pII-e0 probe on the other 7 traces. However, in the second simulation (with 48 common features), the pII-e0/pII-e0 activation was only 1.52 times the prototype/pII-e0 activation, and was overwhelmed by activation of the prototype by the other traces.

Given the success of this simulation in replicating the pattern of effects in Experiment 2, a third simulation was attempted in which 24 of the 48 common features were changed to exemplar-specific features. For example, a subset of features for which all exemplars had the values {+1, -1, +1, -1, +1, -1, +1, -1} in the previous simulation were now set to {+1, -1, -1, -1, -1, -1, -1, -1} for one exemplar, {-1, +1, -1, -1, -1, -1, -1, -1} for a second

[5]For simplicity, it was assumed in the present simulations that the learning parameter L in the model (Hintzman, 1988) was set to 1.0, so that features cannot take on values of 0.

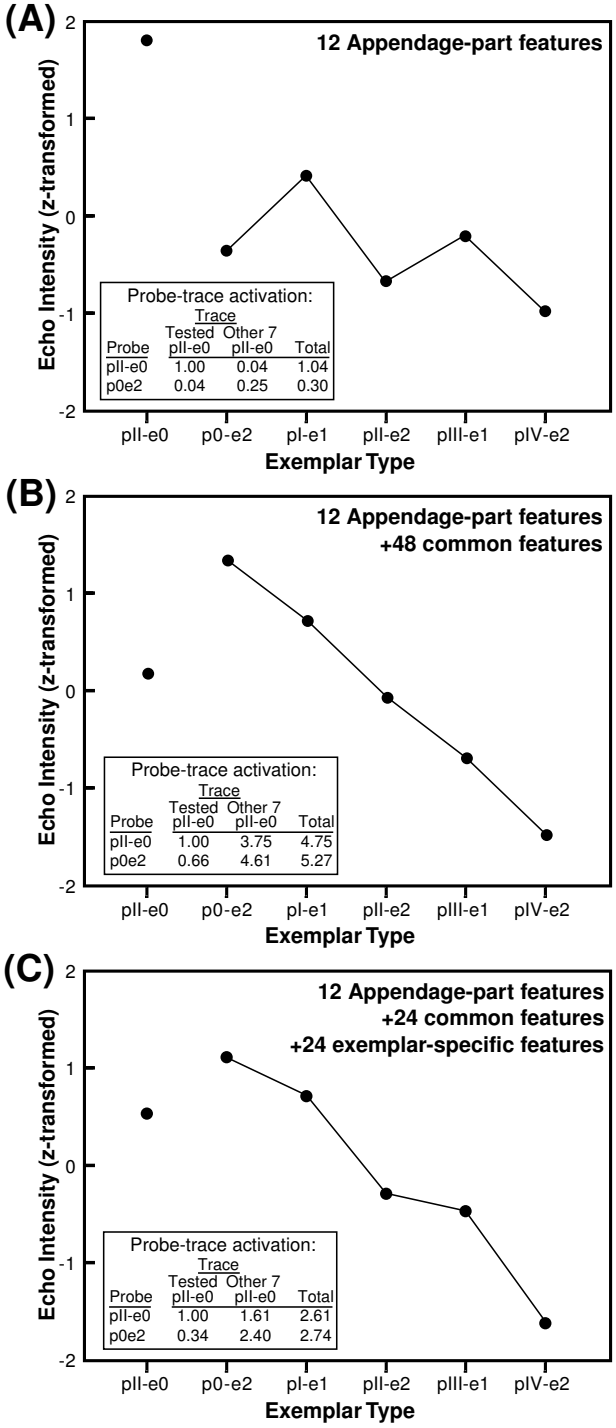

Figure 12: Results of simulations using the MINERVA2 model (Hintzman, 1984, 1986).

exemplar, etc. This change was meant to simulate training on owner-level names in Experiment 3. As shown in Figure 12C, this simulation did produce a decrease in the prototype-advantage effect compared to the previous simulation, just as occurred in Experiment 3 compared to Experiment 2. The slope of the prototype-distance effect also decreased in this simulation compared to the last. However, along with the increase in echo intensity for the trained exemplar came an increase in nearest-neighbor influence, an effect not seen in the data.

Several attempts were made to divorce trained-exemplar intensity from nearest neighbor effects on untrained exemplars, but all were unsuccessful (for example, adding 24 exemplar-specific features to the 48 common features, rather than substituting exemplar-specific for common features, did not alter the pattern of results). This should probably not be surprising: The intertwining of these two factors seems a fundamental property of all multiple-trace memory models with non-linear effects of similarity (e.g., Hintzman, 1986; Kruschke, 1992; Nosofsky, 1991), and this property will be discussed further in connection with the neural-network model introduced below.

Neural-Network Simulations

Exemplar models such as MINERVA2 (Hintzman, 1986) assume that trained category exemplars are encoded separately during learning, and that later categorization performance depends on matching test items to these exemplar-specific representations. Prototype models, on the other hand, assume that the central tendency of a category is abstracted during learning, so that categorization performance can be based on comparisons with category prototypes. Results of the present experiments show strong prototype effects: Test items more similar to species prototypes were responded to more confidently than items more distant from prototypes. However, the simulations reported in the previous section show how exemplar models can also account for this performance pattern. Thus the behavioral predictions of prototype and exemplar models are often equivalent, even though the underlying architectures of the models may be quite different (see Barsalou, 1990 and Estes, 1986 for excellent discussions of this point).

In what could be considered a third class of models, however, even these architectural distinctions become blurred. These are neural-network models that assume that a perceptual input and a behavioral output can be represented as patterns of activation in sets of neuron-like units. Input and output units in these models are joined by a matrix of connections, each of which has a modifiable strength or "weight." Given any input pattern, the connection weights determine an output pattern, and weights can be altered via learning rules to produce various input-output associations.

Knapp and Anderson (1984) and McClelland and Rumelhart (1986) applied such neural-network models to classification paradigms like the one used in the present study. These networks were first "trained" on one set of stimuli, then "tested" on a transfer set with pre-specified relations to the training set and to category prototypes. As in MINERVA2 (and other exemplar theories), these neural-network models did not include any kind of ex-

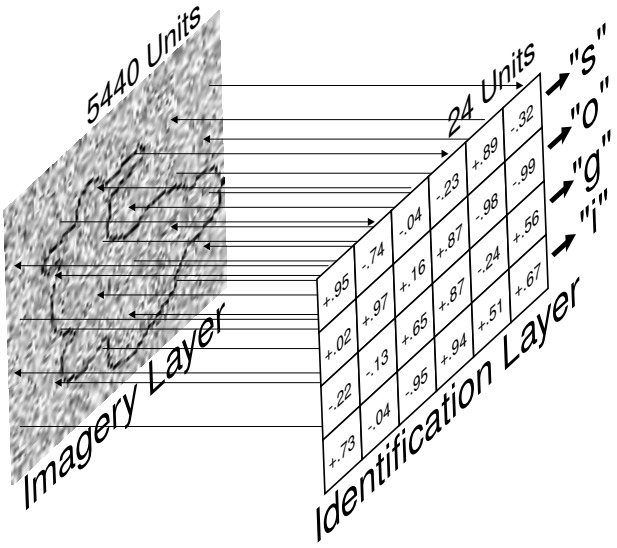

Figure 13: Schematic diagram of the WHOA model.

plicit prototype-abstraction process. On each simulated trial, connection weights were simply modified so that the network became better at classifying the particular training stimulus presented in the trial. However, rather than forming a separate trace for each learned exemplar, these models implicitly encoded all exemplars in the same set of connection weights. Therefore, as in prototype models, information about specific training experiences tended to be lost, since newly-learned exemplars had to overwrite old ones in order to be stored. McClelland and Rumelhart (1986) showed that when trained on exemplars surrounding a central prototype, their network came to demonstrate the same kinds of prototype-distance effects shown in Experiments 2-4 of the present study.

Here, a simple neural-network model is presented that can be trained and tested on images of Fribbles. The first subsection outlines the model, the second subsection reports simulations of Experiments 1-4, and the final subsection discusses why the model works and some of its limitations. In the General Discussion, the model is compared to previously-proposed models of object recognition and categorization.

Architecture of WHOA

The model was dubbed WHOA, for Widrow-Hoff Object Associator. Briefly, WHOA consisted of an *imagery* layer, which coded an image of a Fribble, and an *identification* (ID) layer, which coded arbitrary patterns that could be associated with one or more images (Figure 13). The two layers were linked by independent sets of forward and backward connections. During training, both forward and backward connection weights were altered via a modified Widrow-Hoff learning rule (this rule is also known as the delta rule and the least-mean-squared rule). An analog to Recognition confidence scores was then computed for test items.

Imagery Layer. The imagery layer consisted of 5440 units, each of which represented the intensity of one pixel in an 80 × 68 array. Although the imagery layer could function equivalently given grayscale, color, or black-and-white images, pilot simulations indicated that outlines of the fribble images (examples of which can be seen in Figure 15) produced the most easily interpretable prototype effects. These outlines were thus used in all the simulations reported here.

Identification Layer. The ID layer consisted of 24 units which coded patterns to be associated with images. Unlike imagery-layer patterns, which represented images veridically (i.e. spatiotopically), ID-layer patterns were completely arbitrary (i.e. symbolic). For ease of interpretation, each ID-layer unit was thresholded to -1 or +1, and groups of six units taken to represent letters (a technique borrowed from Anderson, 1995).

Connections and Learning Rule. The two layers were connected by a *forward* set of weights from the imagery to the ID layer and a *backward* set of weights running in the opposite direction. Every imagery-layer unit was connected to every ID-layer unit, and vice versa. Given an input pattern represented by the vector $\mathbf{f}$ on either the imagery or ID layer and the set of connections $\mathbf{A}$ between the two layers, an output pattern on the other layer was computed by matrix multiplication, $\mathbf{Af}$.

Weights were modified during learning by a variant of the Widrow-Hoff rule. Given an input pattern $\mathbf{f}$ and a target pattern $\mathbf{t}$ to be associated with the input, the change in weights $\Delta\mathbf{A}$ was computed by the following formula:

$$\Delta\mathbf{A} = (\mathbf{t} - \iota\mathbf{Af})\mathbf{f}^T,$$

where ι (iota) represents a "generalization constant" which ranges from 0 to 1 and determines how specific the model is in its learning (multiplication of the term in parentheses with the transpose of $\mathbf{f}$, $\mathbf{f}^T$, corresponds to taking the outer product of the two terms). When ι is 1, the model performs full Widrow-Hoff learning, and modifies its weights to try to respond perfectly to the particular input patterns being learned. When ι is 0, the model performs simple Hebbian learning, and essentially ends up responding best to the average of all input patterns learned. For all the simulations reported here, this constant was set at 0.1 (see below for discussion of this parameter).

Learning and Testing Algorithms. During training, the model was given Fribble images and test labels, and both forward and backward weights were modified based on these patterns. Thus when forward connections were modified, $\mathbf{f}$ was the presented image and $\mathbf{t}$ the label to be assigned to the image; for modification of backward connections, $\mathbf{f}$ was the label and $\mathbf{t}$ the image.[6] Labels used in the various simulations are described below.

[6]This learning algorithm corresponds most closely to the Train-View and Train-Press tasks performed in the experiments, in which participants were able to view Fribble exemplars along with their names. More complicated learning algorithms that simulated the Train-Feedback and Train-Test tasks were implemented and reported by Williams (1997). Simulations involving these more elaborate algorithms led to essentially identical patterns of results as the simulations reported here.

To simulate the Recognition test, the outline of the test stimulus was fed through the forward connections and the resulting ID-layer pattern was fed back though the backward connections. Confidence judgments were simulated by taking the cosine of the vectors representing the final *feedback-image* pattern and the presented pattern, $\cos(\mathbf{f}, \mathbf{A}_B(\mathbf{A}_F\mathbf{f}))$, where $\mathbf{f}$ is the presented image, $\mathbf{A}_F$ is the forward-connection matrix, and $\mathbf{A}_B$ is the backward-connection matrix (see Jordan, 1986, for a primer on vector cosines and other linear algebra techniques used here). The vector cosine is appropriate for modelling recognition confidence because it is essentially a measure of "perceptual fluency" (Jacoby, 1983) in WHOA: An item that matches the model's encoded information well will flow smoothly through the system and produce a good match between the final feedback-image and the original image, whereas an item the model is unfamiliar with will produce little activation in the feedback image and thus a small vector cosine between it and the original image.

Simulations

Simulations of Experiments 1, 2, and 3 were performed by training and testing WHOA on outlines of the same objects used in the experiments (unlike in the MINERVA2 simulations, exemplars from all six species were trained and tested together). Before training was begun, connection weights were set to random values between 0 and 1. Weights were then altered according to the training algorithm described above. The number of repetitions on each training exemplar and the labels assigned to each are given in Table 5. In the simulation of Experiment 1 the model was trained to associate the prototype of each species with the species name (sogi, duva, etc.). In the simulation of Experiment 2, the eight pII-e0 exemplars were seen by the model during training, and were also associated with species names. Separate simulations of the Long and Short training protocols of Experiment 2 were run (designated Simulations 2L and 2S), with the proportion of pII-e0 exemplars 1-4 and 5-8 altered accordingly (see Table 2).

For the simulation of Experiment 3, WHOA was trained to associate exemplars 1-4 of the pII-e0 set with names such as "csSG" and "vrFP" (for "Carlos' SOGI" and "Vera's FIPO"), and exemplars 5-8 with names such as "@@SG" and "@@FP" (representing an unnamed SOGI and FIPO). Note that the relative frequencies of 1-4 and 5-8 exemplars in Simulation 3 were identical to those of Simulation 2S, as in the experiments. However, unlike in the experiments, all simulations included the same number of training trials per species. Pilot simulations indicated that it sometimes took this many trials for WHOA to approach asymptotic performance on species-level naming with the pII-e0 training set, and the training results in Tables 2 and 3 make it clear that participants easily reached such performance levels. For this reason and for simplicity in interpreting the simulation results, equal numbers of training trials per species were run in all simulations.

Following training, test objects were processed by the model and responses computed for each object as described above. As in the experiments, the test phase of each simulation was identical. Weights were not modified during testing. Test responses were averaged across items and across 20 runs of each simulation, each with a different random order of training items. The simulation data is shown in Figure 14 along with re-plots of the behavioral data. Values in each plot were scaled so that the difference between the strongest and weakest responses was the same in the human and simulation data.

Table 5: Stimulus repetitions and identification-layer labels in WHOA simulations.

	Simulation							
	1		2L		2S		3	
Ex[a]	N[b]	Label	N	Label	N	Label	N	Label
p	48	sogi	0		0		0	
1	0		6	sogi	8	sogi	8	csSG
2	0		6	sogi	8	sogi	8	vrSG
3	0		6	sogi	8	sogi	8	nnSG
4	0		6	sogi	8	sogi	8	mvSG
5	0		6	sogi	4	sogi	4	@@SG
6	0		6	sogi	4	sogi	4	@@SG
7	0		6	sogi	4	sogi	4	@@SG
8	0		6	sogi	4	sogi	4	@@SG
Total	48		48		48		48	

Note. [a]Ex = exemplar: p = prototype, 1-8 = pII-e0 exemplar number 1-8; [b]N = total number of training trials on this exemplar.

The top-left graph in Figure 14 shows performance in Experiment 1 and Simulation 1, in which training included only the prototypes of each Fribble species. WHOA produced an essentially identical pattern to the human data. The top row in Figure 15 shows WHOA's representation of a prototype as it was built up during training in this simulation. At the top-left of the Figure is the outline of a prototype Fribble that was presented to the model; to the right of this outline are snapshots of the model's feedback-image to this picture just before the second, fourth, sixth, eighth, and tenth learning trials on the object (ID-layer responses and cosines of the original and feedback image vectors are given above each picture). The bottom row in Figure 15 shows what happened when different exemplars of the same species were presented during test. Presentation of the prototype led to a strong response that was highly correlated with the input image, whereas presentation of an item differing by I, II, or III parts from the prototype led to progressively weaker and less well-correlated responses (vector cosines are also given in the Figure). The model produced a feedback-image that had the same shape as the prototype, no matter what exemplar was presented.

The other graphs in Figure 14 plot simulated and human participant data for Experiments 2 and 3, in which participants and WHOA were trained on the pII-e0 training items. Data from the Long and Short training protocols of Experiment 2, which were combined in previous analyses, are plotted separately here. Another change from previous presentations of the data is that exemplars 1-4 and 5-8 from the pII-e0 set are plotted separately (in the two left-most points on the graphs). For the sake of simplicity, results for exemplars 5-8 were not included in previous analyses. They are included now because they

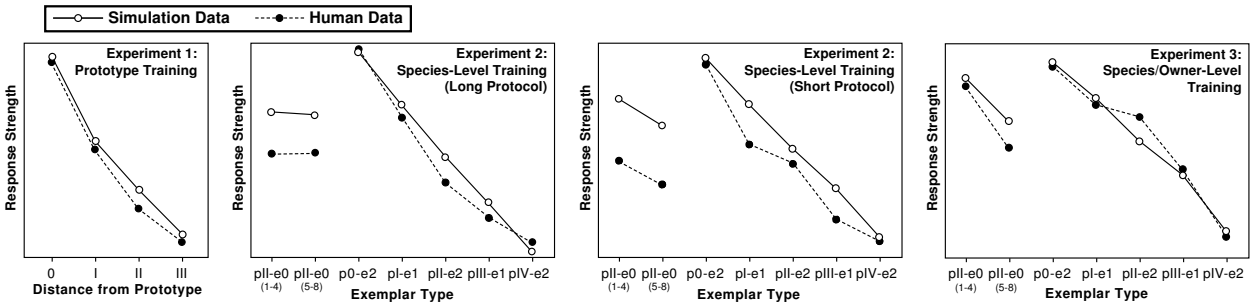

Figure 14: Human and simulation data for Recognition tests in Experiments and Simulations 1-3.

provide a testable prediction for the model: In the Long protocol, exemplars 1-4 and 5-8 were seen equally often, and so should be responded to equally strongly. In the Short protocol, however, exemplars 1-4 were seen twice as often as exemplars 5-8, so the former should be responded to more strongly than the latter. This prediction was borne out in both the model and the human data.

A more striking correspondence between the model and human data concerns the untrained test exemplars. WHOA clearly duplicated the prototype-distance effect, responding most strongly to species prototypes and progressively less strongly to pI-e1, pII-e2, pIII-e1, and pIV-e2 items. The prototype-advantage effect was also replicated in the simulation data: WHOA responded more strongly to previously unseen prototypes than to trained pII-e0 items. The model's test performance in Simulation 2L is represented graphically in Figure 16, which shows responses to a prototype, pII-e0 and pIV-e2 exemplar of a learned species, as well as an exemplar of an unlearned species (numbers beneath the pictures again give vector cosines). As in Simulation 1, the model responded with essentially the same-shaped pattern for all exemplars of a species. Note, however, that unlike in the previous simulation, the feedback-images here were "fuzzy," especially around places where the exemplars differed from each other (e.g., the top of the object). Given an unlearned species, the model produced a very weak pattern that was essentially an average of all six trained species.

As shown in Figure 14, the prototype-advantage and prototype-distance effects were also found in Simulation 3. Furthermore, the sizes of these effects varied from simulation to simulation just as they did from experiment to experiment. As reported in the Discussion of Experiment 3, the prototype/pII-e0 difference was significantly larger in Experiment 2 than in Experiment 3, and the same relationship held in the simulations of these two experiments (differences of .049 and .014 in cosines for Simulations 2 and 3, respectively, $F(1, 58) = 366.5$, $MS_e = .000046$[7]). Likewise, the slope relating distance from the prototype and response strength was steeper in Experiment 2 and its simulation than in Experiment 3 and its simulation (these changing slopes are not represented in Figure 14 because the scale was changed from graph to graph).

[7]For this analysis, simulation run was used as the random variable and results from Simulations 2L and 2S were combined.

Discussion

WHOA successfully modelled the major experimental results found here. Specifically, the model, like human participants, a) responded more strongly to prototypes than to pII-e0 items, even though the former were never seen during training; b) showed graded response strengths to other sets of untrained test exemplars, responding more and more weakly as items differed more and more from prototypes; and c) showed stronger prototype effects (reflected in shallower slopes and smaller differences between prototypes and pII-e0 items) in Simulation 2 than in Simulation 3.

Although WHOA encodes information in a distributed memory store, rather than MINERVA2's (Hintzman, 1986) separate traces, the basis for both models' recognition judgments is the similarity between test items and retained information about individual studied items. As demonstrated above, MINERVA2 was also able to simulate the prototype-advantage and prototype-distance effects in Experiment 2, but made a crucial, incorrect prediction regarding the results of Experiment 3: As the prototype-advantage effect diminished, nearest neighbor effects grew. In other words, manipulations that increased the recognizability of trained items disproportionately increased the recognizability of items similar to trained items. In the present paradigm, this corresponds to a decrease in p0-e2/pI-e1 and pII-e2/pIII-e1 differences and an increase in pI-e1/pII-e2 and pIII-e1/pIV-e2 differences, visible in MINERVA2's simulation results (Figure 12C) but not in the results of Experiment 3 (Figure 10). In WHOA, the size of the prototype-advantage effect was manipulable without introducing nearest-neighbor effects (Figure 14). How was this possible?

Why WHOA works. The first key to answering this question is to note that while both models depend on similarity between studied and test items to determine performance, similarity values in MINERVA2 are cubed before being combined, while the relationship between similarity and performance in WHOA is linear. Thus in MINERVA2, highly recognizable trained items "capture" very similar transfer items, so that the effect of nearest-neighbors on pI-e1 and pIII-e1 items is disproportional to their effect on p0-e2 and pII-e2 items. In WHOA, however, highly recognizable trained items affect the recognizability of all transfer items proportionally.

While a linear similarity function does away with nearest-neighbor effects, it produces a new quandary:

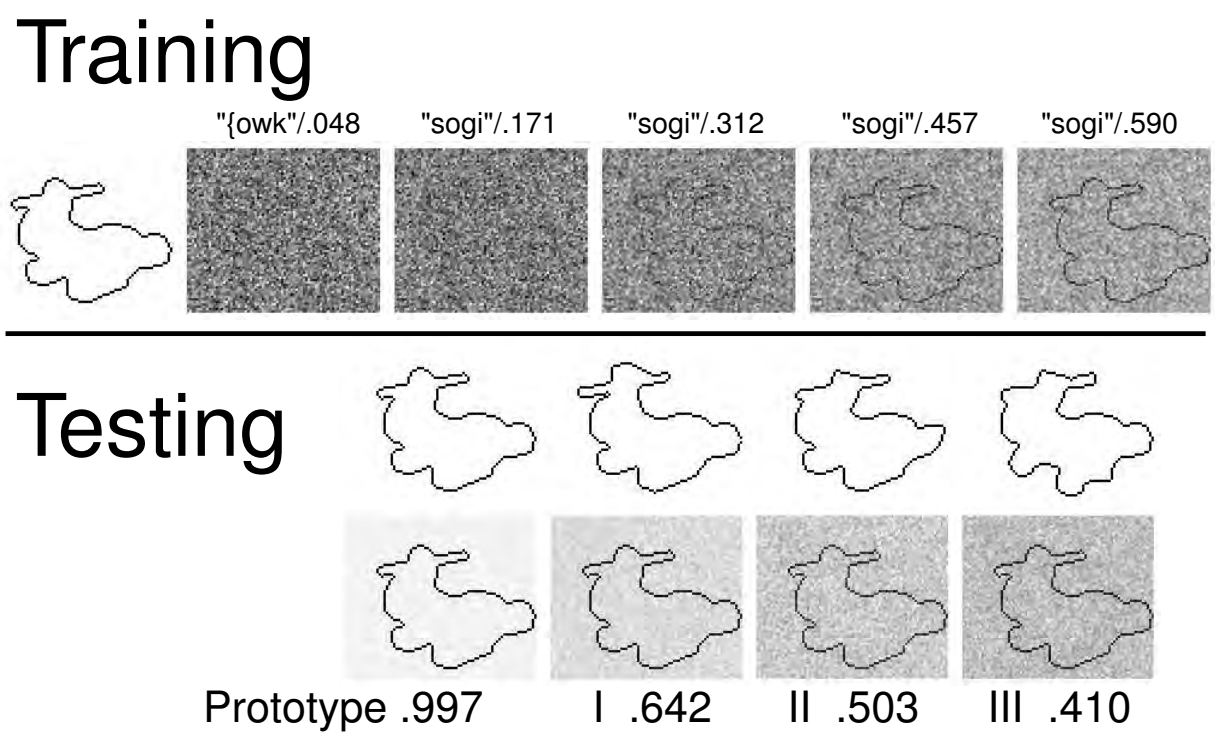

Figure 15: Graphical representation of WHOA's responses to objects in Simulation 1.

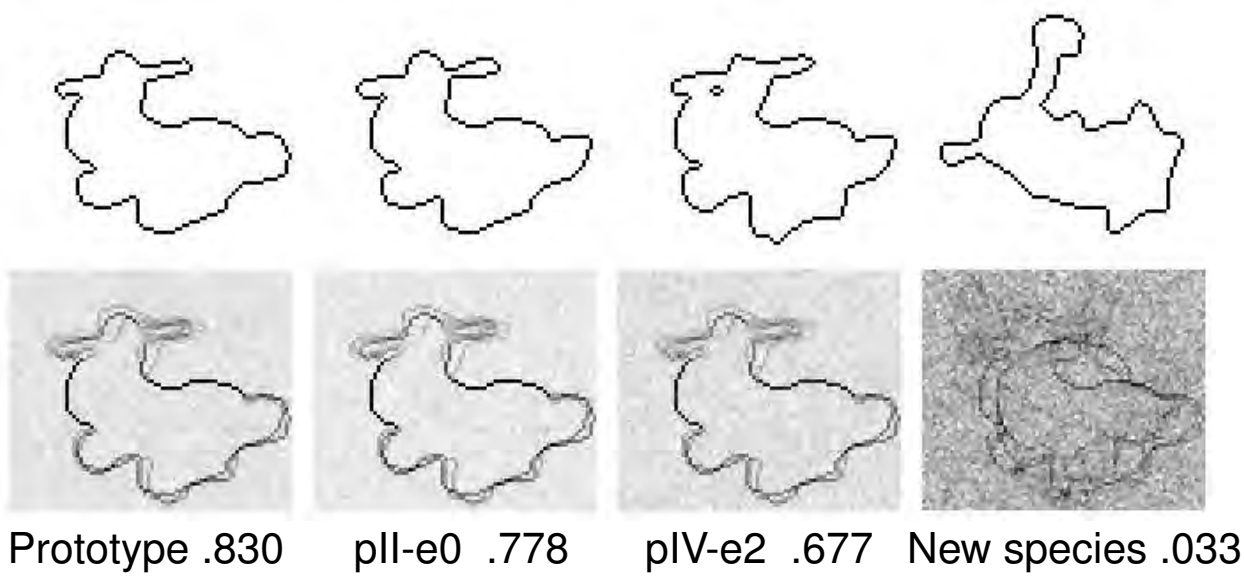

Figure 16: Graphical representation of WHOA's test responses in Simulation 2L to prototype, pII-e0, and pIV-e2 exemplars of a learned Fribble species and an exemplar of an unlearned species.

How can trained (pII-e0) items be recognized with any more confidence than pII-e2 items? (Although the confidence difference between these two exemplar types was small, it was present in each of Experiments 2-4, and combining the results of the three experiments, 37 of 54 participants gave higher confidence ratings to pII-e0 than to pII-e2 exemplars, $z = 2.72$.) Indeed, if MINERVA2 is implemented with a linear (instead of cubic) similarity function, it produces identical echo intensities for these two types of items, since both are equally similar, on average, to the group of trained exemplars as a whole.

To see why WHOA does not exhibit this behavior, consider how the model performed in Simulations 2L and 2S. Here, WHOA learned through modification of forward-connection weights to associate trained items of a species with the same ID-layer pattern, e.g. "sogi." At the same time, the backward-connection weights were modified to produce the trained items on the imagery layer when given "sogi" on the ID layer. Because all the sogi training exemplars were quite similar, and all were quite different from exemplars of other species, the forward connections come to very faithfully produce the "sogi" label in response to any member of the species (even pIV-e2 exemplars produced labels with average cosines of .988 with the correct label). Because all types of test exemplars produced essentially the same label, almost identical feedback-images were generated when these labels were fed through the backward connections, as seen in Figure 16. This feedback-image represented the average of all 8 pII-e0 training items.

Since the prototype of each species was made up of the most common shape in each appendage part socket, it should have been, and usually was, the exemplar that matched this composite feedback-image most closely. Therefore, since the measure of recognition confidence was the cosine between test image and feedback-image, prototypes were "recognized" with the most confidence. Furthermore, exemplars that differed by I, II, III, or IV parts from the prototype tended to match the composite feedback-image progressively less well, and thus received lower cosines and confidence scores.

The quandary, however, still remains: If the cosine was dependent on match with the composite feedback-image, and if this feedback-image was composed of a linear combination of the trained exemplars, then why were pII-e0 items recognized with more confidence than pII-e2 items, which were equally distant from the prototype? One relatively uninteresting possibility would be that the exemplars selected for the pII-e0 set were somehow inherently more "recognizable" (by both humans and WHOA) than the pII-e2 exemplars. This possibility was tested with a pair of simulations in which WHOA was trained either on exemplars 1-4 of the pII-e0 set or on the four pII-e2 exemplars.[8] Results of these simulations were nearly identical: As in previous simulations, prototypes produced higher cosines than trained exemplars, cosines for other exemplars varied by distance from the prototype, and trained

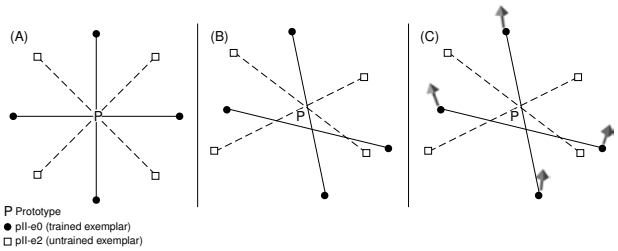

Figure 17: Schematic diagrams of possible similarity spaces created by the WHOA model.

exemplars produced higher cosines than other items that varied by the same distance from the prototype (i.e., when pII-e0 items were trained, they produced higher cosines than pII-e2 items, while when pII-e2 items were trained, these items produced higher cosines).

Further analysis of the network structure and simulation results indicated that if the pII-e0 exemplars of a species had been *exactly* evenly spaced around the prototype exemplar and equally similar to this exemplar, then pII-e0 and pII-e2 items would have received equivalent recognition scores. However, small variations from the idealized similarity space represented in Figure 5B were present in the stimulus set, due to factors such as exactly where appendage parts were placed in each object and the fact that the three different parts that fit into each appendage-part socket were not equally salient in images. Therefore, the average of all training exemplars was less than perfectly correlated with the prototype exemplar, and the composite feedback-image was usually closer on average to pII-e0 items than to pII-e2 items, leading to higher cosines and confidence scores for the former than the latter exemplar type.

Consult Figure 17 to visualize this phenomenon. Figure 17A shows the idealized similarity space, in which all pII-e0 and pII-e2 items are equally similar to and symmetrically arranged around the prototype. Figure 17B shows a more realistic space, in which exemplars are randomly shifted away from their ideal locations (but in which the average distance between exemplars remains constant). The prototype is still closest to the center of the pII-e0 items, which corresponds to the composite feedback-image constructed from these items. However, the average distance from pII-e2 items to this point must be farther than the average distance from pII-e0 items to this point.[9]

Similar considerations explain how WHOA successfully modelled the decrease in the prototype-advantage effect from Experiment 2 to Experiment 3. In Simulation 3, although the second two letters in the ID-layer label still coded species-general information (Table 5), WHOA

[8] The stimulus set was designed so that the pII-e2 items shared the same relationship with other item sets as exemplars 1-4 of the pII-e0 set, i.e. the pII-e2 items surrounded the species prototypes in the same way and shared the same nearest-neighbor relationships with pI-e1, pIII-e1, and pIV-e2 items

[9] As an informal proof of this statement, consider that the center of the space spanned by the exemplars of a set is the point with the shortest average distance from the exemplars. Therefore, any other point in the space is necessarily farther on average from the exemplars. The center of the pII-e0 set corresponds to the composite feedback-image resulting from training on these exemplars. So unless the centers of the pII-e0 and pII-e2 sets coincide (as in Figure 17A), pII-e0 exemplars will be, on average, closer (more similar) to the composite feedback-image than pII-e2 exemplars.

was trained to associate individual items with exemplar-specific information via the first two letters. Thus the model was forced to develop a set of forward-connection weights that produced different ID-layer patterns to very similar imagery-layer patterns (the various pII-e0 exemplars of a given species). This is a much more difficult task than the model was given in Simulations 2L and 2S, where only members of different species had to be discriminated. To the extent that the model was able to construct a successful set of weights, it did so by emphasizing the dimensions on which training exemplars of a species differed from each other.

This emphasis on intra-species variation meant that the representational space WHOA created for the objects was skewed even more by the idiosynchrasies of particular training items. This skewing in turn had two effects on resulting cosine measures: Overall, feedback-images matched test images less well than in Simulations 2L and 2S, but the match for trained items was greater relative to the match for untrained items. Figure 17C attempts to illustrate this phenomenon. Representations of trained exemplars move away from each other, but also move on dimensions (in the Figure, out from the plane of the page) orthogonal to the dimensions that best define other exemplars. (Note that it is difficult to adequately illustrate the intricacies of WHOA's similarity space, since this space is defined by 130,560 dimensions—the number of connections between the imagery and ID layers—rather than the three dimensions represented in the Figure.)

It is important to note that the preceding mechanisms do not depend on the Widrow-Hoff learning component of the model. Additional simulations revealed that the model produced the same basic patterns of results when strict Hebbian learning was implemented by setting the ι parameter to 0. To the extent that Widrow-Hoff learning is allowed to operate (recall that in the main simulations, ι was set to 0.1, where a value of 1.0 would result in full Widrow-Hoff learning), the model's representational space is even more skewed towards coding intra-species variations between exemplars. Thus as ι increases, recognition confidence for trained exemplars rises compared to confidence for untrained exemplars. Interestingly, the relationship between ι and average test image-feedback-image cosines (across both studied and unstudied exemplars) is non-monotonic, increasing as ι moves from 0 to small values, but then decreasing as ι gets larger. Nearest-neighbor effects also emerge with large ι values, for reasons that are not currently well understood. Detailed analysis of the Widrow-Hoff learning component of the WHOA model is an intriguing area for further study.

Limitations of WHOA. Two important discrepancies should be noted between the simulated data produced by WHOA and actual confidence ratings given by participants. First, WHOA consistently overestimated the response strength for trained exemplars (Figure 14). In essence, WHOA's memory for exactly what it saw during training was too good, suggesting that adding noise to the system might bring the model's predictions more into line with the human data. Such noise could correspond, for example, to lapses of attention to certain portions of training exemplars (analagously, features in MINERVA2 were coded probabalistically, rather than perfectly, in Hintzman's, 1986, simulations).

Second, although the overall patterns of simulated and experimental results were quite similar, the correlation across individual items was quite low. To perform this analysis, simulational and experimental data for all 138 test exemplars were first normalized by subtracting the mean score for the item's exemplar type. The correlation between the resulting simulated and experimental data points, $r = .183$, was low enough to strongly suggest that WHOA and participants based their decisions on different sources of information. This should hardly be surprising, since the outline-shape images on which WHOA was trained and tested were extremely impoverished compared to the photorealistically-rendered images shown to participants.

With this fact in mind, it should be emphasized that the structural and procedural assumptions underlying WHOA would function equivalently given almost any type of information about objects. That is, the model will extract the similarity space of a set of training exemplars regardless of the terms in which exemplars are described. In the present simulations, Fribbles were defined as collections of black and white pixels, but objects defined as sets of color pixels, lines, surfaces, three-dimensional volumes, etc. could be handled equally well by the model. Thus the take-home message from these simulations should be that as long as the information extracted from exemplars form a similarity space like that sketched in Figure 17B rather than Figure 17A, an associative neural-net model such as WHOA will successfully duplicate the patterns of effects observed in Experiments 1-4 (see Edelman (in press) for more discussion of this important point).

General Discussion

The experiments reported here reveal an interesting mixture of specificity and generality in object recognition/categorization processes. On the specificity side, when participants were trained in Experiments 1 and 2 to categorize Fribbles by assigning names to each species, they did not simply extract abstract part-structures. Instead, they acquired information about the specific three-dimensional shapes that made up the exemplars of each species, even though such information was incidental to classification performance. On the generality side, however, participants had great difficulty recognizing exactly which exemplars they had seen during training in Experiments 2-4. That is, they seem to have retained information about how specific exemplars relate to their species, but lost much of the information about the specific exemplars themselves. As a result, results showed strong prototype effects, reflected in two performance patterns: Participants were more confident that they had seen species prototypes during training than that they had seen the actual trained exemplars (the prototype-advantage effect), and recognition confidence on other test items was almost perfectly predicted by similarity between test exemplars and species prototypes (the prototype-distance effect).

The most striking aspect of the data reported here is surely the consistent and robust prototype effects in each experiment. However, three additional results have important implications for models of object recognition/categorization. First, similarity between test exemplars and nearest-neighbor trained exemplars was never a significant predictor of recognition performance, de-

spite the fact that Whittlesea (1987), using a very similar stimulus-set design (but with pseudoword stimuli, rather than pictures of three-dimensional objects) found stronger nearest-neighbor effects than prototype effects. Second, the training protocol of Experiment 3, which required participants to distinguish exemplars of the same species as well as distinguishing species from each other, resulted in a significantly reduced prototype-advantage effect and a shallower slope for the prototype-distance effect, but still no detectable sign of nearest-neighbor effects. Third, the training protocol of Experiment 4, which required no Fribble categorization whatsoever, resulted in even stronger prototype effects than found in Experiment 3, demonstrating that explicit acquisition of the Fribble category structure is not a necessary condition for prototype effects to occur.

Following the data-collection phase of the study, simulations were performed to investigate how successful two recognition/categorization models would be in accounting for the experimental results. The MINERVA2 simulations (based on Hintzman, 1986, 1988) indicated that this and other currently-popular exemplar models (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1991) provide a satisfactory account for the basic prototype effects, but predict that increased recognizability of trained exemplars, as observed in Experiment 3 compared to Experiment 2, should be accompanied by increased nearest-neighbor effects, contrary to the empirical data. The neural-network WHOA model represents an alternative both to these exemplar models and to prototype models that posit explicit prototype abstraction processes (Posner & Keele, 1968; Reed, 1972). In WHOA, images of and names for trained objects were implicitly coded in the weights of connections linking two sets of units. Each exemplar was encoded by a separate process, but since all exemplars were stored in the same set of weights, the central tendency (prototype) of each Fribble species ended up producing the strongest response in the model. Simulations using the WHOA model successfully produced decreases in prototype-advantage and prototype-distance effects without noticeable nearest neighbor effects. Implications of the experimental results and the WHOA model for previously-proposed categorization and object recognition models are discussed in greater detail below.

Implications for Exemplar-Based Categorization Models

Most exemplar-based models of categorization (e.g., Medin & Schaffer, 1978; Nosofsky, 1991) posit that a trained exemplar's influence on performance is nonlinearly dependent on its similarity with a given test exemplar. That is, as similarity between a trained and test exemplar increases, the trained exemplar's relevance for categorizing the test exemplar increases disproportionally. Shepard (1987) found this type of relationship to be so pervasive in both the human and animal psychology literatures that he proposed it as a "universal law of psychological science." A corollary to this law is that if one trained exemplar is more similar to a test exemplar than any other trained exemplar, this nearest neighbor exemplar will have an inordinately large influence on categorization/recognition performance.

The sets of test exemplars used in the present experiments were carefully designed to reveal such nonlinear effects of similar trained exemplars on categorization performance. In a stimulus design inspired by that of Whittlesea (1987), every test exemplar differed by exactly zero, one, or two features from its most similar trained exemplar (where features in the present stimuli are defined by the three-dimensional shapes of Fribble appendage parts). If categorization/recognition performance was solely dependent on similarity with the nearest trained exemplar, then trained exemplars (pII-e0 items) would have produced the greatest test response strength, followed by items differing by one part from their nearest trained exemplar (pI-e1 and pIII-e1), followed by items differing by two parts (prototypes, pII-e2, and pIV-e2). Even if performance was only heavily influenced (not completely determined) by nearest neighbors, we should have observed something like the pattern in Figure 7B, reflecting the more reasonable hypothesis (embodied in the theories of Kruschke, 1992, Medin & Schaffer, (1978), and Nosofsky, 1991) that all trained exemplars have some effect on categorization performance, but that highly similar exemplars are disproportionately weighted in decision processes.

Neither of these patterns were found in any of the experiments reported here. Recognition confidence scores were never well-predicted by nearest-neighbor similarity. Instead, performance on untrained test exemplars was consistently and almost perfectly predicted by the similarity of each exemplar to the average trained exemplar, or modal prototype, of the exemplar's species. Note that it cannot be the case that participants simply failed to encode exemplar-specific information during training, for the prototype pattern of effects entails acute knowledge of the particular appendage-part shapes present in training exemplars. Moreover, confidence scores for pII-e0 items were consistently (albeit sometimes only slightly) greater than scores for pII-e2 items, which were equally distant from species prototypes.

Even more vexing for exemplar-based theories is the fact that recognizability of trained exemplars in Experiment 3 rose considerably compared to Experiment 2, yet there was no accompanying increase in nearest-neighbor effects across the two experiments.[10] As demonstrated by the MINERVA2 simulations above, these models predict that an increase in recognizability of pII-e0 items should lead to relative increases in confidence ratings for pI-e1 and pIII-e1 items, compared to prototypes, pII-e2, and pIV-e2 items (Figure 12C).

The lack of such nearest-neighbor effects is especially surprising given that Whittlesea (1987), using a nearly identical stimulus-space design as the one used here, found strong evidence for nearest-neighbor effects, including significantly greater response strength for pIII-e1 than for pII-e2 items. Why were the results of the present experiments so different from those of Whittlesea's? One plausible explanation involves the types of stimuli used in the two studies. Whittlesea used letter-string stimuli, where one of the trained items was NEKAL and its corresponding pII-e2 and pIII-e1 items NOKAP and PEKAL, respectively. The pre-defined modal prototype for this category of letter strings was NOBAL. Letter strings were

[10] An additional experiment reported in Williams (1997) included solely owner-level training, focusing participants even more on intra-species differences between Fribble exemplars, and again failed to produce nearest-neighbor effects.

convenient to use because their features (letters) are well defined and easy to manipulate. However, these pronounceable letter strings must certainly have been interpreted by Whittlesea's participants as words, and the categorical structure of words may be quite different from that of objects. For example, common parts are excellent predictors of similarity between pairs of objects (Tversky & Hemenway, 1984), whereas words cluster into categories by symbolic meanings (FROG–TOAD), not by letter differences (FROG–FROM). Thus Whittlesea's participants may have been biased to treat his letter strings as unitary stimuli, rather than as members of categories.

Indeed, when Whittlesea (1987) forced participants to attend to the category structure of his letter-string stimuli during training (by copying only letters that matched category prototypes, rather than copying entire strings as in his previous experiments), he found no difference between pII-e2 and pIII-e1 items. Whittlesea was able to account for this shift in performance through a model he termed the *episode model*, which included a parameter r that determines the integral of the similarity function. When $r = 1$, the function is linear, meaning that all trained exemplars contribute equally to test performance; as r increases, nearest neighbors contribute disproportionately more than other trained exemplars. From the perspective of the episode model, then, it could be argued that forcing participants to learn species names for Fribbles, as was done in Experiments 1 and 2, led to encoding of trained exemplars with $r = 1$, unlike in most of Whittlesea's experiments. However, the training procedure used in Experiment 3 (learning owner names for specific exemplars) surely should have required participants to encode Fribble exemplars more "integrally" than in the species-level training procedure of Experiment 2. This change did lead to greater recognizability of trained items, yet unlike in Whittlesea's experiments, there was no shift towards nearest-neighbor patterns of effects on the Recognition test.

Three-dimensional objects, much more than words, can usually be validly categorized on the basis of physical similarity, at least at the entry level (Cutzu & Tarr, 1997). Most objects that have wings and a beak are birds, and in the present stimulus set, all objects that have a long main body and two horizontal "feet" are SOGIs. This is true even if we are learning that some birds are robins and others cardinals or that one SOGI is owned by Nancy and another by Carlos. Thus we may be so used to dealing with three-dimensional objects in terms of category distinctions that we encode objects with relation to other category members (i.e., with $r = 1$) even when not explicitly asked to do so.

Implications for Prototype-Based Categorization Models

Even if this conjecture is correct, it does not explain how or in what form entry-level category information is encoded. Perhaps the simplest explanation is that "abstract ideas" (Posner & Keele, 1968) are developed during training and used during categorization task performance. That is, the central tendencies (prototypes) of categories are extracted from training exemplars, and test exemplars are compared to prototypes to determine category membership.

Although intuitively appealing, prototype theories raise many subtle but difficult questions. For example, how would a system responsible for prototype abstraction "know" how general a prototype should be, given that only a subset of the possible exemplars of a category are presented to the system during training? The present stimulus set offers an excellent example of this problem. What should the prototype for the SOGI category look like? Since the Fribble species were designed to differ in their abstract part structures, the best prototype would be an abstract part description (top-left picture in Figure 3). Such a representation would require the least amount of detail to be encoded, but would describe all SOGI exemplars equally well and be equally dissimilar to all exemplars of other species. If participants had formed abstract part structures during training and used them at test, they should have shown no confidence differences at all between test sets. Thus the fact that every experiment revealed significant effects of exemplar type is a strong indicator that participants did not acquire abstract part descriptions during training (or, at least, they did not use such descriptions even if they were formed).

Instead, participants responded as if they had formed representations including the most common shape for each appendage part of each species (the pattern shown in Figure 7A). Thus if a prototype extraction mechanism was at work, it seems to have chosen the modal prototypes to represent the Fribble species. This brings us to an even more vexing issue. Participants always saw items singly, and every training item included two prototypical appendage parts and two non-prototypical parts. How would the extraction system build a prototype from imperfect exemplars that are presented only one at a time?

At least two possible solutions to this problem have been proposed in the categorization literature. The first is embodied in the "feature list" approach to category learning, in which the prototype-extraction system keeps a running tally of the features present in the exemplars of each category, and treats the collection of features with the highest counts as the prototype (Franks & Bransford, 1971; Hayes-Roth & Hayes-Roth, 1977; Reitman & Bower, 1973). This approach provides a straightforward mechanism for prototype creation and updating. However, it only works well when category features are well defined, as they are in the geometric-form stimuli used by Medin and Schaffer (1978) and others. While Fribble parts themselves are well-defined, the *relations* between parts (e.g. the fact that the SOGI's two "feet" are below its main body) are critical in distinguishing species from each other, and it is unclear how one could succinctly represent all the possible relations between object parts in a feature list. Other problems with purely feature-based approaches to categorization are discussed by Schyns, Goldstone, and Thibaut (in press).

A second possible abstraction mechanism would require storing individual trained exemplars in memory, at least temporarily, so that the modal features could be determined from collections of trained exemplars. Individual-exemplar information could then be discarded (or at least largely disregarded) once the prototypes were formed. This approach introduces a host of new questions. How long do the exemplars need to be retained before the prototype is suitable for use in categorization? How should the system deal with multiple instances of the same ex-

emplar? How, exactly, is the modal prototype computed? Moreover, as pointed out at the beginning of this paper, some information about individual instances must be permanently retained if we are to be able to identify special cases of categories (Mike's house). Indeed, even though participants in Experiments 2 and 4 were not asked to remember any specific instances of the Fribble species, they were at least slightly more confident that they had seen trained exemplars than would be predicted by these exemplars' similarity to modal prototypes (see Hintzman, 1986, for further consideration of the problems facing prototype extraction models).

Thus a satisfactory explanation of prototype effects such as those found in the present study should provide both for permanent storage of at least some aspects of trained exemplars, along with a seamless transition from pure exemplar representations to central-tendency representations. Furthermore, some account must be made for why the strength of the prototype-distance and prototype-advantage effects were manipulable via the different training procedures used in Experiments 2-4.

WHOA as a Categorization Model

Neural-network models such as WHOA are well-suited to accomplish these goals. WHOA distinguishes trained exemplars from untrained exemplars (although imperfectly, just as human participants do), but at the same time produces powerful prototype effects. Furthermore, these prototype effects fall directly out of the combination of learning rule and data representation implemented in the model—no special process is needed to explicitly abstract prototypes. Thus the model provides the seamless transition from exemplar to prototype representations alluded to above: When WHOA knows only one exemplar for a Fribble species, that exemplar is the prototype; when WHOA learns a second and third exemplar for the species, the prototype is automatically updated to reflect the newly-learned information (Knapp & Anderson's (1984) and McClelland & Rumelhart's (1986) models, as well as a rather different neural-network model proposed by Schyns, 1991, also demonstrate these desirable properties).

Importantly, WHOA was also able to simulate the changing magnitudes of the prototype-distance and prototype-advantage effects between Experiments 2 and 3. This behavior again fell out of the model because of the data representations used in the simulations. When all exemplars of a species were associated with exactly the same name, the slope was quite steep, and when the number of common letters in exemplar names decreased to two, the slope decreased as well. Whittlesea's (1987) exemplar model could also account for shifting slopes, by manipulating the integration parameter r. As demonstrated above, changing slopes are also modelable in MINERVA2 (Hintzman, 1986) via somewhat different mechanisms. However, both these manipulations also alter the sensitivity of test responses to similarity with nearest-neighbor trained exemplars, an effect that was not observed in the human data. In WHOA, the slope of the prototype-distance effect decreases without a concomitant increase in sensitivity to nearest neighbors.[11]

[11] Interestingly, manipulation of the generalization constant ι in WHOA has a very similar effect to manipulation of r in the episode model, or to manipulation of the similarity gradient q in Kruschke's (1992) ALCOVE model. As noted above, though, WHOA was able to simulate all experiments quite nicely while keeping ι at a fixed value throughout all simulations.

Implications for Object Recognition Models

The present study also has implications for models of object recognition. The most widely known and accepted object recognition theory is Biederman's (1987) recognition-by-components (RBC) model and its neural network implementation, JIM (Hummel & Biederman, 1992). Although JIM, like WHOA, has a neural network architecture, the former is a much more complex model, consisting of some seven layers of units that perform a number of highly specific computations. Essentially, JIM takes as input a line drawing of an object, produces a structural description of the object, and responds via a layer of grandmother-cell units that each uniquely code one structural description.

An important distinction is made in JIM between structural descriptions and representations of objects, because any one entry-level object category may require many structural descriptions to fully represent all of its exemplars. Thus structural descriptions in JIM do not correspond precisely to the abstract part-descriptions discussed here; rather, JIM codes the specific three-dimensional shape of each part of an object. This coding scheme is necessary for the model to distinguish, for example, axes and mallets, since both objects consist of a long thin part BELOW a short thicker part.

In the context of the present experiments, the fact that JIM does not code abstract part descriptions means that it does not make the incorrect prediction that all Fribbles of a learned species will be responded to with equal strength. However, as it is presented in Hummel and Biederman (1992), JIM does not generalize from a known structural description of an entry-level object category to a second, unstudied structural description of the same category. This means that in Experiment 1, JIM might not know that any object other than the prototype (the sole trained instance in this experiment) was a SOGI. It should be possible to build a mechanism into JIM such that when a previously unknown structural description is encountered, it attempts a guess based on the structural descriptions it does know. But even with this modification, JIM would predict that trained items should be recognized with greater facility than all untrained items, even prototypes; this prediction is falsified in each of Experiments 2-4. Such behavior is a fundamental property of the theory's principle that structural descriptions be coded discretely, rather than probabilistically. That is, RBC codes the many instantiations of an entry-level category by coding each of the possible structural descriptions of the category separately, rather than by developing a single representation of the category that matches all exemplars to a greater or lesser extent.

In its dealings with categories as a whole, then, RBC is quite different from prototype theories. The correspondence pointed out in the General Introduction was between RBC's and prototype theories' handling of individual exemplars, where both develop a single represen-

tation to handle multiple encounters with an individual object (e.g., from different viewpoints). In other words, both RBC and prototype theories posit abstract representations (i.e., representations which are divorced from specific encoding episodes), but the unit of abstraction is different in the two cases. Prototypes abstract over all instances of a category, no matter how diverse the category may be, whereas RBC only abstracts over instances of a structural description, and the limits of when structural descriptions change are clearly defined. Therefore, as a model of categorization, the present study demonstrates that RBC is somewhat impoverished, despite its popularity as a model of object recognition (see also Tarr & Bülthoff, 1995).

Other models have attempted to account for object recognition phenomena through neural network architectures that work with two-dimensional, exemplar-specific representations, rather than three-dimensional structural descriptions. The earlier-proposed models (Edelman & Weinshall, 1991; Poggio & Edelman, 1990) dealt solely with recognition of single objects from different views, and as such did not directly address the issues of categorization brought up by the present experiments and the WHOA model. However, since these models utilize Hebbian learning rules, they would probably capture some aspects of the prototype-formation phenomenon observed in WHOA. More recent models from this group (Bricolo, Poggio, & Logothetis, in press; Edelman, 1995) deal more explicitly with categorization issues, and an informative future exercise will be to evaluate the similarities and differences between WHOA and these models.

WHOA as an Object Recognition Model

Like Edelman's (1995) *chorus of prototypes* model, WHOA is a model of both object recognition and object categorization. That is, the model can learn both to recognize objects at the individual level (Vera's SOGI) and at a categorical level (SOGI). Furthermore, both levels are handled through exactly the same representations and processes (see Gauthier, Williams, Tarr, and Tanaka, in press, for another example of WHOA's individual and categorical recognition abilities, on a completely different object set). In terms of parsimony, this is a great advantage over models that assume fundamentally different representations and/or processes for evaluating categorical and individual-instance information. Such a two-process assumption is made explicitly in many categorization models (e.g., Elio & Anderson, 1981; Posner & Keele, 1968) and implicitly in structural description models such as RBC (Biederman, 1987), since multiple instances with the same structural description will have to be kept track of by separate mechanisms. Another advantage of WHOA over other object recognition models is that it can account for performance shifts over changes in encoding tasks (Experiments 2-4); most if not all current object recognition models are insensitive to encoding context manipulations.

However, the WHOA model does not deal explicitly with the pre-eminent issue in object recognition research, namely identification of objects from different viewpoints. In its current form, WHOA could easily process and encode images of one or more Fribbles from different viewpoints, and it will be interesting to see how well the model handles such input. Regardless of how the present version of the model performs, its capacities should be greatly increased by further extensions, especially making the model a dynamic system, for example by adding a "brain-state-in-a-box" (BSB; Anderson, 1995) mechanism on top of the model's current architecture. BSB models are designed to drive imperfect inputs towards stable states; in the present context, this behavior could correspond to mapping known objects from novel viewpoints onto representations of the objects from already-encoded viewpoints.

Implications for Recognition Memory Models

Although the paradigm employed here was derived primarily from the categorization literature, the test task used was old/new recognition, so it may be asked how the results relate to the recognition memory literature. One point of contact involves "false recognition" effects found in word-list memory experiments (Deese, 1959; Roediger & McDermott, 1995). For example, Roediger and McDermott (1995) found that after study of a 12-item list including such words as THREAD, THIMBLE, and THORN, participants were more likely to say they had seen the word NEEDLE (probability .76), which was strongly associated to the words on the list but not actually presented, than to say that they had seen actually-presented items (.72). This finding has obvious parallels with the prototype-advantage effects found in Experiments 2-4.

However, the false recognition effects found in the present study (as well as those in some previous categorization experiments, e.g. Franks & Branford (1971), and Homa et al. (1993)) were for stimuli that participants had never seen prior to the start of the experiment. Thus the present results would be difficult to attribute to *implicit associative responses* (Roediger & McDermott, 1995; Underwood, 1965), in which study of words such as THREAD and THIMBLE activates the previously-existing memory trace for the word NEEDLE. Here, no representation of the prototype of a Fribble species existed in participants' minds prior to the start of training—the first time participants saw the prototype was on the recognition test, when they judged they had seen it before with greater confidence than actually-studied exemplars.

It is possible that participants explicitly built representations of species prototypes during training, and that the high confidence given to prototypes at test was due to source monitoring errors (Mather, Henkel, & Johnson, in press). In other words, an explicitly abstracted prototype could constitute a qualitatively similar memory trace as the traces formed in response to the studied pII-e0 exemplars, and participants may have failed to correctly attribute prototype traces to extraction processes. Furthermore, it is easy to imagine that in Experiments 2 and 3, when species categorization was part of training, prototype traces would actually be activated on every trial, making them stronger than pII-e0 traces and resulting in the strong prototype-advantage effect observed. Such an explanation is less convincing in the context of Experiment 4, however. No species categorization was required of participants in this experiment, so an explicit abstraction/source monitoring explanation should predict a smaller prototype-advantage effect in this experiment than in Experiment 3, where categorization was required during training. In fact, the effect was larger in Exper-

iment 4 than 3. Another problem with this explanation is that in addition to prototypes, pI-e1 items were also judged OLD with higher confidence than studied pII-e0 items in Experiments 2 and 4. It is unlikely that pI-e1 items would have been explicitly abstracted during training.

These results could perhaps be better explained by *global matching models* of recognition memory (Clark & Gronlund, 1996), of which Hintzman's (1988) MINERVA2 is a prime example. The common denominator of these models is that test items are used as cues to activate memory traces of studied events, and recognition decisions are based on the total amount of activation engendered by the test item. In the context of the present paradigm, Fribble exemplars could be stored as quadruple associations (i.e., ABCD, where each letter represents a three-dimensional volume that can vary from exemplar to exemplar within a species). Recognition of a trained exemplar would proceed by presenting ABCD as a test probe, whereas recognition of a novel test exemplar that varied by two parts from this trained exemplar would involve presenting ABEF as a test probe. This was essentially the strategy used in the MINERVA2 simulations presented above.

Results of these simulations indicated that the nonlinear relationship between probe-trace similarity and activation (echo intensity) posited in MINERVA2 led this model to an incorrect prediction: As the recognizability of trained exemplars rose from Experiment 2 to Experiment 3, MINERVA2 predicted that nearest-neighbor effects on untrained exemplars should also have risen. The SAM model (Gillund & Shiffrin, 1984) would seem to make the same prediction, due to its assumption that cues are combined multiplicatively (a similar assumption is made in the context model of Medin & Schaffer, 1978). However, a detailed evaluation of SAM's handling of the present stimulus-set design is beyond the scope of this paper.

An alternative both to these individual-trace models and to models positing explicit formation of a discrete prototype trace is provided by distributed memory models, such as TODAM (Murdock, 1983), CHARM (Eich, 1982), the Matrix model (Pike, 1984), and WHOA (see Clark & Gronlund, 1996, for descriptions of all the global matching models mentioned here). As Knapp and Anderson (1984, p. 617) put it, "the fundamental assumption behind the distributed memory approach is that remembered items share many or all of the same storage elements, so that one cannot properly point to a single memory trace." While distributed memory models give rise naturally to prototype effects, there is nothing special about the prototypes themselves, except that they happen to be at the center of the similarity space spanned by pII-e0 items and thus share the maximum number of "storage elements" with learned patterns. In this framework, the difference between prototype (p0-e2) and pI-e1 exemplars is not qualitative, as an explicit abstraction model would imply, but quantitative, as indicated in the data. Whether or not the global matching models cited above would be able to adequately predict the present results is, again, beyond the scope of the present enterprise. However, the consistent lack of nearest-neighbor effects in the experimental data make distributed memory models more attractive than discrete-trace models.

Conclusion

No extant model of categorization or object recognition alone provides a satisfactory account for the total of the present set of data. What is needed is a model that specifically addresses the concerns of both the object recognition and categorization literatures. The WHOA model represents a start in this direction, taking actual pictures as input and producing category or individual names as output. Like other neural-network categorization models (Knapp & Anderson, 1984; McClelland & Rumelhart, 1986), WHOA combines desirable aspects from both prototype/structural description models, in that generalization to unencountered exemplars proceeds naturally and efficiently, and exemplar/image-based models, in that a single set of processes and representations is employed to identify stimuli at the individual and categorical levels. Future research and modeling efforts should continue to explore the feasibility and benefits of combining individual and categorical representations in the same processing system.

References

Anderson, J. A. (1995). *An introduction to neural networks.* Cambridge, MA: MIT Press.

Barsalou, L. W. (1990). On the indistinguishability of exemplar memory and abstraction in category representation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition* (Vol. 3, p. 61-88). Hillsdale, NJ: Erlbaum.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115-147.

Bricolo, E., Poggio, T., & Logothetis, N. K. (in press). 3D object recognition: A model of view-tuned neurons. In *Advances in neural information processing systems 9.* California: Morgan Kaufmann.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin and Review, 3*(1), 37-60.

Corballis, M. C., Zbrodoff, N. J., Shetzer, L. I., & Butler, P. B. (1978). Decisions about identity and orientation of rotated letters and digits. *Memory and Cognition, 6*, 98-107.

Cutzu, F., & Tarr, M. J. (1997). The representation of three-dimensional object similarity in human vision. In *SPIE proceedings from electronic imaging: Human vision and electronic imaging II.* San Jose, CA.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17-22.

Edelman, S. (1995). Representation, similarity, and the chorus of prototypes. *Minds and Machines, 5*(1), 45-68.

Edelman, S. (in press). Representation is representation of similarity. *Behavioral and Brain Sciences.*

Edelman, S., & Weinshall, D. (1991). A self-organizing multiple-view representation of 3D objects. *Biological Cybernetics, 64*, 209-219.

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review, 89*, 627-661.

Elio, R., & Anderson, J. R. (1981). The effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 397-417.

Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology, 18*, 397-417.

Franks, J. J., & Bransford, J. D. (1971). Abstraction of visual patterns. *Journal of Experimental Psychology, 90*, 64-74.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (in press). Training "Greeble" experts: A framework for studying expert object recognition processes. *Vision Research.*

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1-65.

Hayes-Roth, B., & Hayes-Roth, F. (1977). Concept learning and the recognition and classification of exemplars. *Journal of Verbal Learning and Verbal Behavior, 16*, 321-338.

Hintzman, D. L. (1986). "Schema" abstraction in a multiple-trace memory model. *Psychological Review, 93*(4), 411-428.

Hintzman, D. L. (1988). Jundgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*(4), 528-551.

Homa, D., Goldhardt, B., Burruel-Homa, L., & Smith, J. C. (1993). Influence of manipulated category knowledge on prototype classification and recognition. *Memory and Cognition, 21*, 529-538.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 418-439.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99*(3), 480-517.

Jacoby, L. L. (1983). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 21-38.

Jolicoeur, P., Gluck, M., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology, 16*, 243-275.

Jordan, M. I. (1986). An introduction to linear algebra in parallel distributed processing. In *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, p. 365-422). Cambridge, MA: MIT Press.

Knapp, A. G., & Anderson, J. A. (1984). Theory of categorization based on distributed memory storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 616-637.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22-44.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin and Review, 1*, 476-490.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Philosophical Transactions of the Royal Society of London B, 200*, 269-294.

Mather, M., Henkel, L. A., & Johnson, M. K. (in press). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory and Cognition.*

McClelland, J. L., & Rumelhart, D. E. (1986). A distributed model of human learning and memory. In *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, p. 170-215). Cambridge, MA: MIT Press.

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

Murdock, B. B. (1983). A distributedmemory model for serial-order information. *Psychological Review, 90*, 316-338.

Nosofsky, R. M. (1991). Test of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance, 17*(1), 3-27.

Pike, R. (1984). A comparison of convolution and matrix distributed memory systems. *Psychological Review, 91*, 281-294.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature, 343*, 263-266.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353-363.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83*, 304-308.

Reed, S. K. (1972). Pattern recognition and categoriation. *Cognitive Psychology, 3*, 382-407.

Reitman, J. S., & Bower, G. H. (1973). Storage and later recognition of exemplars of concepts. *Cognitive Psychology, 4*, 194-206.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803-814.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382-439.

Schacter, D. L., Cooper, L. A., & Delaney, S. M. (1990). Implicit memory for unfamiliar objects depends on access to structural descriptions. *Journal of Experimental Psychology: General, 119*, 5-24.

Schyns, P. G. (1991). A modular neural network model of concept acquisition. *Cognitive Science, 15*, 461-508.

Schyns, P. G., Goldstone, R. L., & Thibaut, J. (in press). The development of features in object concepts. *Behavioral and Brain Sciences.*

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237*, 1317-1323.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23*, 457-482.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study of the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review, 2*(1), 55-82.

Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon-structural-descriptions or by multiple-views? *Journal of Experimental Psychology: Human Perception and Performance, 21*(6), 1494-1505.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation dependence in shape recognition. *Cognitive Psychology, 21*, 233-282.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General, 113*, 169-193.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition, 32*, 193-254.

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology, 70*, 122-129.

Whittlesea, B. W. A. (1987). Preservation of specific experiences in the representation of general knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 3-17.

Whittlesea, B. W. A., Brooks, L. R., & Westcott, C. (1994). After the learning is over: Factors controlling the selective application of general and particular knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 259-274.

Williams, P. (1997). *Prototypes, exemplars, and object recognition.* Unpublished PhD thesis, Yale University.