Strategy Development and Learning Differences in Supervised and Unsupervised Categorization

Journal:	*Memory & Cognition*
Manuscript ID:	MC-RES-07-295.R2
Manuscript Type:	Resubmitted Manuscript
Date Submitted by the Author:	n/a
Complete List of Authors:	Colreavy, Erin; University of Western Australia, School of Psychology Lewandowsky, Stephan; University of Western Australia, School of Psychology
Keywords:	unsupervised categorization, supervised categorization, category learning

THE UNIVERSITY OF
WESTERN AUSTRALIA

FACULTY OF
Life and Physical
Sciences

7 December 2007

Dr Brian Ross, Editor

Memory & Cognition

Re: MC-RES-07-295.R1

Dear Dr Ross:

Thank you very much for your action letter and acceptance of our manuscript "Strategy Development and Learning Differences in Supervised and Unsupervised Categorization" pending revision. Your comments were both interesting and very useful in our current revision, and we hope we have addressed the two main issues of paradigm and manuscript length to your approval.

Concerning paradigm clarification, we thank you for pointing out the concerns that readers may have. We have softened our conclusions in response, and we hope that our paper now evokes further thoughts and discussion rather than strong disagreement in readers.

Concerning length of the manuscript, we have endeavoured to remove all excess information. We have managed to condense the manuscript into a more succinct 24 pages of text (with six figures), which is equivalent in length to the last reviewed submission, and 15% shorter than the accepted revision. We hope that this fulfils the length requirements expressed in your action letter.

We are grateful for your acceptance of this paper and we thank you again for your comments and suggestions throughout this review process.

Sincerely,

Erin Colreavy

Strategy Development and Learning Differences in

Supervised and Unsupervised Categorization

Erin Colreavy and Stephan Lewandowsky

University of Western Australia

Running head: Unsupervised Categorization

Address correspondence to: Erin Colreavy

School of Psychology

University of Western Australia

Crawley, W.A. 6009, AUSTRALIA

colree01@student.uwa.edu.au

Abstract

The processes that determine unsupervised categorization, the task of classifying stimuli without guidance or feedback, are poorly understood. Two experiments examined the emergence and plasticity of unsupervised strategies using perceptual stimuli that varied along two separable dimensions. In the first experiment, participants either classified stimuli into any two categories of their choice or learned identical classifications by supervised categorization. Irrespective of the complexity of classification, supervised and unsupervised learning rates differed little when both modes of learning were maximally comparable. The second experiment examined the plasticity of unsupervised classifications by introducing novel stimuli half-way through training. Irrespective of whether or not people altered their strategies, they responded to novel stimuli in a gradual manner. The gradual and continuous evolution and adaptation of strategies suggests that unsupervised categorization involves true learning which shares many properties of supervised category learning. We additionally show that the choice of unsupervised strategy cannot be predicted from the properties of early learning trials but is best understood as a function of the initial distribution of dimensional attention.

Strategy Development and Learning Differences in

Supervised and Unsupervised Categorization

To date, most laboratory research on categorization has focused on supervised category

learning; that is, acquisition of a pre-determined category structure, such as when medical

students learn to classify diseases by their symptoms. However, we must often categorize objects

in the absence of any "right" or "wrong" answer. For example, how people divide their e-mails

into "interesting" versus "spam" is completely personal, with no external error correction or pre-

determined classification (Gureckis & Love, 2003). Some theoreticians have considered this self-

directed or unsupervised categorization to be a fundamental component of all forms of

categorization (e.g., Schyns, 1991). Notwithstanding, unsupervised categorization has attracted

relatively little attention and the underlying processes are not well understood (Love, Medin, &

Gureckis, 2004). Likewise, although several models of unsupervised categorization have been

proposed (Clapper & Bower, 1991, 1994; Fried & Holyoak, 1984; Love et al., 2004; Pothos &

Chater, 2002), their explanatory power does not rival those for supervised categorization.

Several important questions about unsupervised categorization remain unresolved, such as

why people decide on one classification rather than another and how exactly they acquire their

chosen strategy in the absence of feedback. Do people randomly pick a classification strategy or

do they *learn* to categorize, comparable to supervised categorization? If unsupervised

classification involves learning, how does it differ from supervised learning? How responsive is

unsupervised categorization to alterations in the environment? We present two experiments that

sought answers to these basic questions about unsupervised categorization.

Concerning the choice of classification strategy, a common hypothesis is that people seek to

create "simple" structures. For example, using colored discs as stimuli, Imai (1966) found that

people preferred a smaller number of categories to a multitude of groups; that they preferred

categories with equal numbers of stimuli; and that they tended to focus on one dimension along

which to divide stimuli. The latter observation was confirmed by Ashby, Queller, and Berretty (1999), who showed that people could classify perceptual stimuli based on either one of two dimensions (length and orientation of lines), but were unable to discover an experimenter-designated classification involving variation along both dimensions.

More recently, Pothos and Chater (2002) proposed that unsupervised categorization was governed by simplicity, as defined by the shortest possible coding—in terms of bits in the Minimum Description Length (MDL)—of the inequalities between all pairwise relationships among stimuli. Because within-category similarities are, by near-axiomatic assumption (Pothos & Chater, 2005), expected to be greater than between-category similarities, the creation of categories reduces the overall number of inequalities that need to be specified, thus potentially providing a more efficient description. In several experiments, Pothos and Chater (2002) showed that people generally but not exclusively preferred simple classifications to more complex ones.

The idea of simplicity is also embodied in a current computational model of supervised and unsupervised categorization, known as SUSTAIN (Supervised and Unsupervised STratified Adaptive Incremental Network; Love et al., 2004). According to SUSTAIN, categorization begins with all stimuli assigned to the same group, and categories are added only as necessary in response to a "surprising" event. In unsupervised categorization, a surprising event is a new stimulus that mismatches current knowledge.

Other existing proposals are mute about how people acquire classification strategies. For example, Pothos and Chater's (2002) simplicity principle is oblivious to the dimensions that define stimuli (such as size or color of a rectangle) because it is exclusively concerned with inequalities between pairwise similarities. Moreover, the simplicity principle—and much of the supporting research—only applies to situations in which the number of categories is unconstrained and in which all stimuli are visible simultaneously. Simultaneous presentation precludes any examination of learning because the order in which people process stimuli remains

unknown. Sequential presentation, by contrast, not only permits assessment of learning over time but by implication also permits a comparison of supervised and unsupervised classification.

Existing studies using sequential presentation have answered a limited set of questions. Haubensak (1992) showed that categorical judgments about stimuli were made early on, after only a few trials. Ashby et al. (1999) likewise noted that uni-dimensional strategies were rapidly acquired during a single block of 80 unsupervised trials, although their analysis provided only a coarse measure of learning. Moreover, neither Haubensak nor Ashby et al. compared supervised to unsupervised learning. We know of only one such comparison: Love (2002) presented participants with four-dimensional stimuli. Supervised participants were trained on one of four classifications, ranging in complexity from uni-dimensional, to two dimensional, to no regularity (i.e., instances had to be memorized). Unsupervised participants observed the stimuli or rated them for pleasantness, followed by a recognition test. Complexity affected supervised learning in the expected manner, whereas for the unsupervised participants, the classification in which all dimensions were required turned out to be easier to learn than the simpler, two-dimensional strategy. However, the study permits only limited conclusions because for the unsupervised participants, learning was measured by a recognition test rather than categorization responses.

The paradigm used in our studies differs from precedents by permitting the simultaneous examination of people's *choice* of classification strategy while also tracing the trial-by-trial *acquisition* of the chosen strategy. To date, examinations of learning have usually precluded simultaneous examination of the choice of strategy: Experimenters typically designed stimuli with an underlying but hidden rule and tested people's ability to acquire that experimenter-designated "correct" rule (e.g., Ashby et al., 1999, Billman & Knutson, 1996). When people were found *not* to use the hidden rule favored by the experimenter, as in the multi-dimensional condition of Ashby et al. (1999), participants were deemed unable to learn that strategy. We argue against this view because it fails to acknowledge the possibility that although people may "fail" to classify the stimuli in ways anticipated or preferred by the experimenter, they may nonetheless

learn a perfectly legitimate alternative strategy. Our point is best illustrated by a study by anthropologist Joe Glick (1968; cited in Triandis & Berry, 1980; see also Cronbach, 1990), who asked Liberian Kpelle farmers to categorize stimuli involving food items and utensils. A "mature" categorization, according to Western ideals, divides the stimuli into the two categories "food items" and "utensils." Contrary to that expectation, Kpelle farmers paired each food stimulus with its appropriate utensil, thus grouping the stimuli in pairs, an arrangement considered "immature" by some (cf. Triandis & Berry, 1980). We argue instead that the only difference between these classifications is that one was expected by the experimenter whereas the other one was not.

In our studies, we did not constrain people's choice of classification and encouraged a wide range of choices by using a fairly homogeneous set of stimuli. We nonetheless examined learning of a person's chosen strategy across trials, without reference to an experimenter-designated but hidden "correct" solution. Moreover, to place that learning into context, we contrast the dynamics of the emergence of unsupervised categorization strategies to supervised category learning.

Our first experiment compared unsupervised and supervised categorization under identical conditions. Stimuli varied along two perceptually separable dimensions, one of which was binary (i.e., consisting of two levels) and the other quinternary (i.e., 5 levels). Stimuli were presented sequentially one at a time and remained visible until a classification judgment had been made. In the unsupervised condition, people were free to classify stimuli into any two categories of their choosing; in the (yoked) supervised condition, people were trained on the observed unsupervised classifications using corrective feedback. In our second experiment, we focused on unsupervised learning and examined how people adapted their strategies in response to novel stimuli.

Based on previous results (e.g., Pothos & Chater, 2002), we expected some heterogeneity among people's unsupervised strategies, albeit with a preference for unidimensional classifications (e.g. Ashby et al., 1999; Imai, 1966; Milton & Wills, 2004). Accordingly, another focus of our analyses was on identifying predictors for unsupervised strategies based on

properties of the initial training stimuli. To maximize statistical power, this analysis involves data

from both studies and is reported after the experiments.

<div align="center">Experiment 1</div>

Experiment 1 compared unsupervised and supervised categorization across two yoked

groups of participants. Because supervised training involved the classifications used by the

unsupervised group, all unsupervised participants were tested first.

<div align="center">*Method*</div>

Apparatus and Stimuli

The experiment was controlled by a Windows PC using a Matlab program written with the

aid of the Psychophysics Toolbox (Brainard, 1997; Pelli, 1991).

The stimuli were squares of different sizes and shadings of gray similar to those used by

Goldstone (1994). The stimuli comprised a 5×5 matrix of squares formed by orthogonally

varying size and brightness. The square sizes were 3, 6, 9, 12, and 15 cm and the levels of

brightness (or luminance; measured with a photometer) were 0.6, 1.5, 3.6, 8.1 and 14.5 candelas.

Figure 1 shows the multi-dimensional scaling (MDS) solution for the stimuli obtained from a

group of 10 participants who provided pairwise similarity ratings (using KYST, Kruskal, Young,

& Seery, 1978, with a city-block distance metric, Stress = .16). As expected, the MDS solution

mirrored the stimulus layout almost perfectly, with the columns in the figure corresponding to

squares of different sizes and the rows corresponding to different brightnesses. The MDS solution

confirmed thatseparations between stimuli were roughly equidistant along both dimensions.

Each participant was trained on 10 unique stimuli in one of two combinations, either

involving two different sizes fully-crossed with 5 brightness values (size-binary) or vice versa

(brightness-binary). In the size-binary combination, only the smallest two squares were used, each

combined with all 5 levels of brightness (see dashed box in Figure 1). Conversely, the brightness-

binary combination used the two darkest squares and all 5 sizes (dotted box in Figure 1). People

within each condition were randomly assigned to the size-binary or brightness-binary

combination. (Initial analyses showed that the size-binary and brightness-binary combinations did

not differ and all results are therefore reported collapsed across this counter-balancing measure.)

Participants

One hundred and eight undergraduates (67 female, mean age 19.2 years) at the University of

Western Australia participated in return for course credit. Fifty one participants were assigned to

the unsupervised condition and 55 participants to the supervised condition (a further two

supervised participants did not exceed chance during training and were omitted from analysis.)

Participants in the supervised condition were randomly assigned to be trained on one of the

unsupervised strategies, with between 9 and 11 participants in each group (discussed later).

Procedure

In both conditions, there were 10 blocks of training, each involving a new randomized

sequence of the 10 stimuli. On each trial, a stimulus was presented in the center of the screen.

Unsupervised participants were instructed to create two categories that seemed natural and

intuitive and were told that there was no correct or incorrect answer. Responses were made using

the 'a' and 'l' keys to assign each stimulus to one of the two categories, arbitrarily labeled

"Category A" and "Category B." No feedback was presented and trials were separated by 1 s.

The supervised condition was identical except that each response was followed by corrective

feedback. The feedback consisted of the word "Correct" or "Wrong" appearing in the centre of

the screen for 1 s. The nature of the feedback depended on the category to which a given stimulus

was assigned; this assignment, in turn, depended on the training group—yoked to an unsupervised

classification strategy—to which the participant had been assigned.

Results and Discussion

Unsupervised Strategy Identification

To permit unambiguous strategy identification, stimuli were deemed to belong to the

category to which they were assigned most often (or last, if the proportions were equal).[1] Because

unsupervised category labels were arbitrary, we reverse-coded responses where necessary to

ensure consistent identification of classifications. For example, all participants who used the

binary dimension were considered to have acquired the same strategy irrespective of whether the

larger or the smaller (or brighter or darker) squares were called "A."

Of the 48 unsupervised participants retained for analysis, 17 (35%) elected to categorize

along the binary dimension. Another 17 participants (35%) categorized along the quinternary

dimension, using any one of the four possible ways in which this dimension could be divided.

The majority of these quinternary categorizers attempted roughly even divisions, placing the

boundary between groups of two and three dimensional values. A further 5 participants (10%)

chose multi-dimensional boundaries. These categorizations were so similar to the quinternary

classifications (differing only by the assignment of a single stimulus) thatall remaining analyses

considered them together with the quinternary categorizers. None of the conclusions are altered if

multi-dimensional categorizers are considered separately.[2]

The three strategies just discussed—i.e., binary, quinternary, and multi-dimensional—

formed the basis for training in the supervised condition (with quinternary training involving a

classification that placed the category boundary after the first two dimensional values.)

Supervised and Unsupervised Learning

Learning over trials. In supervised categorization, accuracy of responding provides

an unambiguous measure of performance. In unsupervised categorization, our favored analogue

measure consisted of the consistency of a person's responses, during the course of training, with

their ultimate categorization strategy. We traced learning by considering the proportion of

responses in trials 1-3, 3-5, 8-10, and 18-20 that matched the strategies that were either ultimately

adopted (for the unsupervised participants) or trained (supervised participants). Using these

measures, we present two perspectives on the course of learning in the two conditions.

Figure 2 shows the mean matching proportions for binary (top panel) and quinternary

(bottom panel) categorizers. Each panel in the figure shows clear differences between supervised

and unsupervised performance. After only three trials, the unsupervised categorizers match their

ultimate strategy far more consistently than their supervised counterparts, suggesting that people

settle on their preferred unsupervised classification very early on, whereas supervised learning

occurs considerably more slowly. This difference appears somewhat more marked for the binary

categorizers. In confirmation, a 2 (Condition: unsupervised vs. supervised) $\times$ 2 (Strategy: binary

vs. quinternary) $\times$ 4 (Trial: 3, 5, 10, 20) between-within ANOVA revealed small to moderate

significant main effects of trial, $F(3, 276) = 27.61$, $p < .05$, $\eta^2 = .23$, $MSE = .05$ and condition,

$F(1, 92) = 10.52$, $p < .05$, $\eta^2 = .10$, $MSE = .13$. The interaction between Strategy and Trial

approached significance, $F(3, 276) = 2.12$, $p = .10$, $\eta^2 = .02$, $MSE = .05$.

One unique aspect of unsupervised categorization is that, unlike in feedback-driven learning,

a person's first response can never be "incorrect." That is, irrespective of subsequent stimuli and

irrespective of a person's final strategy, the first response provides a "correct" category label and

an anchor point for subsequent learning. In supervised learning, by contrast, the first response can

serve this function only if it happens to be correct. To maximize comparability between

conditions, we therefore also examined performance of only those supervised participants who,

by chance, provided a correct response on the first trial ($N = 8$ out of 20 for binary and $N = 22$ out

of 39 for quinternary). A 2 (Condition: unsupervised vs. supervised—first response correct) $\times$ 2

(Strategy) $\times$ 4 (Trial) between-within ANOVA revealed a main effect of trial only, $F(3,189) =$

15.15, $p < .05$, $\eta^2 = .19$, $MSE = .04$, with none of the other effects being significant. The analysis

suggests that the differences between supervised and unsupervised learning are largely eliminated

when the role of the first response is equalized between conditions.

The implications of this second analysis can be illuminated further by plotting performance

over trials with respect to both principal strategies (i.e., binary and quinternary). That is, for each

set of trials (1-3, 3-5, 8-10 and 18-20, as before), the proportions of responses that matched each

participant's own strategy (whether trained by feedback or adopted without supervision) were

plotted against the proportion of responses that matched the alternative strategy. For quinternary

categorizers, responses were matched to their chosen strategy and the alternative binary strategy;

for binary categorizers, matching-proportions were averaged across all four possible quinternary

alternative strategies before being plotted against the binary matching-proportions. Figure 3

displays the progress of learning from the first set of trials (large plotting symbols) along the path

indicated by arrows until asymptote is reached (trials 18-20, last plotting symbol in sequence of

arrows). The top panel displays the path diagrams for the binary categorizers, and the bottom

panel shows the paths for the quinternary categorizers. In both panels, the paths originate fairly

centrally and then migrate towards a matching proportion of 1.0 for the chosen strategy, with a

concomitant decrease of the match to the alternative strategy. The figure shows that when only

the supervised participants whose first response was correct are considered, their learning occurs

at roughly the same rate and follows a nearly identical path as unsupervised strategy acquisition.

Category Switches. The path diagrams demonstrate how participants' categorizing becomes

more aligned with a particular strategy over time. However, the path diagrams cannot illuminate

whether people initially flip back and forth between alternative strategies (e.g., alternately using

the binary and quinternary dimension) or whether they use a single dimension throughout but

learn to use it more consistently. This question can be resolved by examining responses to

dimensional changes between successive stimuli. This analysis considered the proportion of times

that people placed two consecutive stimuli into different categories as a function of the pair-wise

difference in the quinternary dimension value. The proportion of category switches was computed

for each condition and each strategy separately, using three "snapshots" at trials 1-5, trials 8-12

and trials 16-20. These observed proportions can then be compared to "predicted" proportions:

For unsupervised participants, the predicted proportions are those that would be obtained if that

participant categorized the particular trial sequence in perfect accordance with the final strategy.

For supervised participants, the predicted proportions are derived from the reinforced stimulus sequences for each participant.[3] Figure 4 shows the proportion of category switches for the binary (top row of panels) and quinternary (bottom row of panels) categorizers, with each column of panels corresponding to one of the three trial periods. Within each panel, we show the results for unsupervised (solid lines) and supervised (dashed lines) participants.

For binary categorizers (top panels), any difference in quinternary value should instigate a category shift *only* if it is accompanied by a difference in binary value; hence the predicted switch proportion (represented by dotted lines) is 0.5 regardless of the quinternary value difference. At the outset of learning, shift proportions were quite haphazard, but by trials 16-20 they closely followed the predicted proportions. For binary categorizers the elevated switch proportion—and its increase over learning—when two successive stimuli share the same quinternary value (i.e., a difference of 0) is noteworthy. Because presentation was randomized within each block, repetition of a quinternary value was more likely to be accompanied by the *other* binary value; hence binary participants were increasingly more likely to switch categories when two successive stimuli had the same quinternary value. Turning to the quinternary categorizers, the unsupervised participants conformed to the predicted switching proportions from very early on.

A 2 (Condition) × 2 (Strategy) × 3 (Trial period) × 5 (Quinternary value difference) between-within ANOVA found a significant main effect of quinternary value difference, $F(4, 72) = 6.44$, $p < .05$, $\eta^2 = .26$, $MSE = .04$. A significant interaction between value difference and condition was small, $F(4, 72) = 2.71$, $p < .05$, $\eta^2 = .13$, $MSE = .04$, and between value difference and categorization strategy was large, $F(4, 72) = 29.34$, $p < .05$, $\eta^2 = .62$, $MSE = .04$.

In summary, the fact that the first five trials exhibit proportions that are already beginning to resemble the predicted switching curves strongly suggests that participants have selected their relevant dimension by this point. This suggests that participants do not waver between

dimensions early in the task, but gradually learn more about the dimension that is relevant to their final strategy.

We recognize that unsupervised and supervised learning are not entirely comparable in our study. To begin with, unsupervised participants are choosing their own strategy, whereas the supervised participants are learning a strategy that is not necessarily of their own preference. In addition, supervised participants have the advantage of feedback which by definition is absent in the unsupervised task. Nonetheless, we find it noteworthy that performance is remarkably similar between the two modes of learning when other differences are controlled to the extent possible.

Learning vs. guessing. One might argue that unsupervised strategy acquisition in our study did not represent true "learning" but instead reflected mere guessing. Specifically, people might have randomly picked a dimension along which to categorize during the first few trials and may have used that dimension for allremain ing stimuli without learning anything about the category structure.[4] We argue against this possibility for several reasons.

First, one implication of this guessing view is that if people randomly pick the binary dimension, their performance should be virtually perfect within a few trials because memory for only two dimensional values would permit flawless classification. The data provide little support for this possibility, because neither the Strategy × Trial × Condition nor the Strategy × Condition interactions that would be expected on the basisof rapid and flawless acquisition of the binary strategy reached conventional levels of significance.

Second, a guessing-based explanation begs the question as to how people classify stimuli along the quinternary dimension. That is, even if the dimension were picked at random, this by itself does not obviate the need for a boundary placement; we are unable to conceive how this could be achieved without some form of learning of the stimuli and differentiation between them.

Third, if unsupervised strategy acquisition involves a guessing process that is distinct from supervised (or any other form of) learning, then it is difficult to explain why the two modes of

learning became indistinguishable in the first-response-correct analysis. This analysis is more

compatible with the view that both modes involve learning.

Finally, a guessing-based account would likely predict that people remain indifferent to any

changes in the stimuli once they have settled on a pattern of responding. By contrast, if

unsupervised categorization involved a mechanism similar to supervised learning, people would

remain sensitive to changes in the stimuli throughout learning. Experiment 2 examined this issue.

Experiment 2

In supervised categorization, it is well known thatpeople can adapt to changes to the

category space. For example, Kruschke (1996) showed successful adaptation to an unannounced

shift in category structure, regardless of whether this shift involved a simple reversal of category

labels or a more difficult strategy involving a previously irrelevant dimension (see Macho, 1997,

for related results). Shifts in supervised categorization can also be obtained by providing an

explicit hint partway through training, without any alteration of the stimuli (e.g., Kalish,

Lewandowsky, & Davies, 2005; Lewandowsky, Kalish, & Griffiths, 2000). Overall, it is clear

thatsupervised learning remains sensitive to a variety of external influences.

By contrast, little is known about the plasticity of unsupervised categorization, although

there are reasons to expect people to respond to changes in the stimuli. For example, SUSTAIN

(Love et al., 2004) postulates that people recruit a new category cluster on the basis of "surprise";

that is, the extent to which a novel stimulus differs from others seen up to that point. However,

unlike in supervised categorization, we know of no experimental investigation of the plasticity of

unsupervised learning; Experiment 2 provides a first step towards examination of this issue.

Method

Experiment 2 consisted of two phases of unsupervised trials. The first phase was a partial

replication of Experiment 1, using only the size-binary combination of stimuli. In the second

phase, additional training stimuli were recruited from the full category space in Figure 1, with the

selection of novel stimuli differing across between-subjects conditions. In all conditions, the

formerly binary size dimension was augmented by adding either one or two novel sizes of stimuli

(each involving all 5 values of the quinternary dimension). The novel sizes were either contiguous

with the sizes used in Phase 1, or skipped a size in order to create a dimensional discontinuity,

with number and contiguity of the new sizes being fully crossed to yield four conditions.

These conditions were chosen to provide a broad range of opportunities for boundary

adjustment and followed two suggestive precedents. First, Imai (1966) found that people

preferred to create groups of equal size; this preference cannot be maintained with an original

(binary) category boundary when the binary dimension turns into a ternary or quaternary

dimension in Phase 2. People may thus be encouraged to adjust their boundary in order to re-

establish an equilibrium between categories. Second, Maddox, Filoteo, Lauritzen, Connally, and

Hejl (2005) reported that categorization was affected by whether or not stimuli were contiguous

along a dimension. Although their study involved supervised training, it nonetheless suggests that

breaking the continuity of a dimension in Phase 2 may encourage participants to move their

boundary to the highly-conspicuous discontinuity.

Participants

A total of 98 participants (22 male; mean age = 20.2 years) were randomly assigned to one

of four conditions: the extra-column condition ($N = 22$), the balance condition ($N = 26$), the gap

condition ($N = 26$), and the balanced-gap condition ($N = 24$).

Stimuli and Procedure

Phase 1 of this study was an exact replication of the unsupervised size-binary combination

(i.e., involving the stimuli identified by dashed lines in Figure 1) in Experiment 1. Participants

categorized unique 10 stimuli, formed by the orthogonal combination of two sizes and 5 levels of

brightness. Each stimulus was shown 10 times, in a completely randomized order.

Phase 2 involved novel stimuli of additional sizes, whose selection differed between

conditions. In the *extra-column* condition, stimuli of the third-largest size were added, to create a

stimulus set with 15 squares varying in 3 contiguous sizes × 5 levels of brightness. The *balance*

condition added the next two larger squares, yielding 4 contiguous sizes × 5 levels of brightness

for a total of 20 stimuli. The *gap* and *balanced-gap* conditions, respectively, added squares of one

or two additional sizes; however, they omitted the middle size. Thus the gap condition used 3

sizes (columns 1, 2, and 4 in Figure 1) × 5 levels of brightness, and the balanced-gap condition

involved 4 sizes (all but the middle column in Figure 1) × 5 levels of brightness.

Phases were separated by a self-timed break, and with the exception of the additional stimuli,

the procedure of Phase 2 was identical to that of Phase 1. Owing to the differences in the number

of stimuli across conditions, the extra-column and gap conditions involved 150 trials (in

completely random order) in Phase 2, and the balance and balanced-gap condition involved 200.

Participants were not alerted to the change of stimuli in Phase 2.

Results and Discussion

Strategy Identification

Categorization strategies were determined in the same manner as in Experiment 1 and were

considered separately for each phase and condition. Two participants were removed from

consideration due to ambiguous categorizations.[5] Two distinct strategies were identified for Phase

1. As shown in Table 1, most participants used a binary strategy, with the remainder using the

quinternary dimension.[6] (In Table 1, dimensional values 1-5 correspond to rows or columns of

stimuli in Figure 1, with the category boundary denoted by a "|".)

During Phase 2, the inclusion of additional sizes allowed people to divide this initially-

binary dimension in a number of ways. About half of the participants who used the binary

category boundary continued to do so in Phase 2 ("1 | others" in Table 1). One exception to this

pattern involved the balanced-gap condition, in which fewer participants retained their original

binary strategy. Most people in that condition moved their category boundary to the gap ("1 2 |

others" in Table 1), suggesting that the combination of discontinuity and the ability to equalize

category sizes encouraged people to revise their strategies.

It is also worth noting that the brightness (quinternary) dimension attracted additional

participants in Phase 2, suggesting that the introduction of new stimuli caused some people to

switch from one dimension to the other. Overall, many participants in all conditions were

sensitive to the addition of new information and adapted their strategies in response.

Collapsing across conditions and classifying strategy shifts between the two phases, we

identified four broad types of participants: Binary-binary categorizers ($N = 24$) continued to

apply their original boundary between the smallest two sizes in Phase 2. Binary-adjust

categorizers ($N = 30$) continued to classify by size in Phase 2, but moved their category boundary

from its original position to (or near) the center point of the now-extended dimension (i.e.,

between the second- and third-largest squares in the extra-column and balance conditions or

between the second- and fourth-largest squares in the gap and balanced-gap conditions). Binary-

quinternary categorizers ($N = 12$) moved from the initially-binary dimension to one of the four

possible quinternary strategies. Finally, quinternary-quinternary categorizers ($N = 23$) utilized a

quinternary strategy in both Phases 1 and 2. Although the category boundary used by quinternary-

quinternary participants did occasionally move, we do not consider these few ($N = 5$) instances

separately. A further 7 participants (including three who placed their category boundary between

the largest two sizes in the balanced and balanced-gap conditions) did not fall into any of these

four types and were excluded from further analyses.

The remaining analyses focus on the differences between these distinct classes of

participants rather than the original experimental conditions.

Learning

Phase 1. As in Experiment 1, unsupervised learning was assessed by comparing performance

on trials 1-3, 3-5, 8-10 and 18-20 to the strategy that each participant ultimately adopted. Figure 5

(top panel) shows the mean matching proportions for binary and quinternary categorizers. Similar

to Experiment 1, both types of categorizers rapidly settled onto their final strategy and were

performing at near-ceiling levels by trial 20. A 2 (Strategy) × 4 (Trial) between-within ANOVA

(condition was not included as a factor because the conditions were identical during Phase 1)

yielded a significant main effect of trial, $F(3, 282) = 22.23$, $p < .05$, $\eta^2 = .19$, $MSE = .04$, and an

interaction between trial and strategy, $F(3, 282) = 7.06$, $p < .05$, $MSE = .04$. The interaction

suggests that binary and quinternary categorizers learned the strategies at different rates; however,

the size of this effect was quite small ($\eta^2 = .07$).

Phase 2. People's adaptation to the new stimuli was traced by analyzing the overlap between

responses during the usual sets of trials to each individual's original Phase 1 strategy (middle

panel of Figure 5) and evolving Phase 2 strategy (bottom panel). The bottom panel shows that all

Phase 2 strategies were acquired quite gradually, with the accompanying trade-off in terms of

unlearning of the initial strategies shown in the middle panel. The unlearning data were entered

into a 4 (Type of strategy shift) × 4 (Trial) between-within ANOVA, which revealed significant

main effects of trial, $F(3,255) = 3.68$, $p < .05$, $\eta^2 = .04$, MSE = .06, and strategy, $F(3,85) = 4.08$,

$p < .05$, $\eta^2 = .13$, MSE = .26, but no interaction between the two variables.

The apparent unlearning of binary-binary categorizers may appear surprising in light of the

fact that their category boundary did not change; its existence is due to the fact that some

participants ($N = 5$) reversed their category labels in Phase 2. This also occurred with a small

number of quinternary-quinternary categorizers, who adjusted their boundary along their favored

dimension, resulting in a similar dip in their performance. It is worth noting that the binary-

quinternary participants not only exhibited the greatest extent of unlearning of their Phase 1

strategy, but also performed worse than the other groups in Phase 2. This suggests that it takes

considerable learning to re-establish a boundary along a different dimension from the one used in

Phase 1. The binary-adjust categorizers, by contrast, were performing almost at the same level as

people who did not change strategy. Changing from one dimension to another appears more

detrimental to category learning than an adjustment of a boundary along the same dimension.

In confirmation, a 4 (Type of strategy shift) × 4 (Trial) between-within ANOVA on the data in the bottom panel of Figure 5 found a significant main effect of strategy, $F(3,85) = 25.13$, $p < .05$, $\eta^2 = .47$, MSE = .06, indicating that performance with novel stimuli varies depending on whether the original Phase 1 strategy is retained or abandoned. No other effects were significant.[7]

Transfer Performance

The first cycle of stimulus presentation in Phase 2 constitutes a transfer test of people's knowledge after acquisition of their initial strategy and before they have adapted to the novel stimuli (ignoring the learning that may take place during the first presentation of the novel stimuli). Figure 6 shows the overall proportion of "A" responses to each of the size and brightness value combinations (in each case considering only those participants who were shown that particular stimulus as dictated by condition), collapsed across conditions and shown separately for people who used a binary (top panel) or quinternary strategy (bottom panel) in Phase 1.

Both panels show smooth and monotonic generalization gradients along the dimension used in Phase 1, suggesting that both groups begin Phase 2 by applying their original boundary in a manner that is consonant with generalization profiles observed in supervised categorization; nonetheless, during further training, many of these participants changed their categories and thus demonstrably learned more about the stimulus structure.

Overall, the learning data converge on the conclusion that unsupervised categorization strategies evolve quickly but nonetheless gradually, and that they retain considerable plasticity in response to changes in the environment. On the basis of the present data and the reasoning offered in connection with Experiment 1, we therefore reject the idea that unsupervised responding is based on an initial random guess and we do not consider this option further. We argue instead that unsupervised strategy acquisition involves category learning in much the same way as supervised training. We return to a discussion of the differences and similarities between the two modes of learning after an analysis of potential determinants of unsupervised categorization strategies.

Determinants of Unsupervised Categorization Strategy

One goal of this article was to determine whether people's choice of strategy could be predicted on the basis of properties of the initial training stimuli. To maximize power, we analyzed Experiment 1 (unsupervised condition) and Experiment 2 (Phase 1) together.

Analysis of Potential Predictors

When presented with the first of a novel set of stimuli for classification, very little information is available on which to base a response. Notwithstanding these impoverished circumstances, the first stimulus may affect classification on subsequent trials. For example, if the first stimulus involves the extremes of a dimension, such as an obviously very large or a very small square, this may draw the participant's attention to that dimension (i.e., size) notwithstanding the absence of any comparison stimuli. In consequence, people may be more likely to rely on a dimension whose values on the first (few) trial(s) were extreme. We therefore considered the quinternary value of the first stimulus as a potential predictor of strategy.[8]

Table 2 shows the proportion of times that each quinternary dimension value (1=smallest or brightest; 5=largest or darkest) appeared on the first trial. The distribution deviated little from chance (.2) and also did not differ much between the two principal strategies, $\chi^2(4, N = 137) = 5.71, p > .10$. Strategy is thus unrelated to the quinternary value of the first stimulus.

Alternatively, people may be sensitive to the dimensional contrast between successive stimuli (i.e., a large stimulus followed by a small one). A number of theories acknowledge the importance of the discrepancy between a current stimulus and its predecessor. For example, Stewart, Brown, and Chater (2002) presented a Memory And Contrast (MAC) model, which states that if a stimulus is highly similar to the one preceding it, it is more likely to be placed in the same category. Conversely, if a stimulus is very different to the preceding one, a reversal in assignment is likely. The MAC model has successfully accounted for sequence effects in supervised category learning.

Accordingly, for each pair of stimuli during the first 5 trials, we counted the number of

dimensions that changed (either 1 or 2; 0 being impossible because no stimulus was repeated

during those trials) and independently determined the city-block distance between stimuli. When

aggregated across the first 5 trials, neither measure differed between binary and quinternary

categorizers: In terms of dimensional changes, the means for binary and quinternary categorizers,

respectively, were 5.77 and 5.86, $t(136) = -.51$. In terms of city-block distance, the means for

binary and quinternary categorizers were 9.08 and 9.06 respectively, $t(136) = .05$, $p > .10$. (The

same outcome was obtained in separate analyses that considered only the first 2, 3, or 4 trials.)

We conclude that inter-stimulus contrast also does not predict people's strategy.

Finally, we entered all potential predictors (first-trial distribution, number of pairwise

dimension changes, and pairwise city-block distance) into a logistic regression model with

strategy (binary vs. quinternary) as the discrete response variable. None of the variables were

found to be significant, largest $\chi^2(3, N = 137) = 1.44$, $p > .10$, confirming that categorization

strategies are unlikely to be determined by the properties of the stimulus sequence.

This result is surprising in light of the clear evidence that people had at least partially

acquired their chosen strategy after as few as 3-5 trials,at which poin t they had seen no more than

half of all to-be-classified stimuli. Thus, the data present us with a paradox: On the one hand,

people rapidly settle on their preferred strategy, but on the other hand, the attributes of those

stimuli seemingly have no bearing on strategy choice. How can this paradox be resolved? One

possibility is that we simply overlooked the crucial predictive stimulus attribute. This possibility

cannot be ruled out, although we are not aware of other attributes which prior theorizing would

identify as relevant. We therefore favor another possibility, which is that people instantly divide

their attention between the various stimulus dimensions and that this initial distribution of

attention, rather than properties of the stimuli, guides subsequent classifications.

Allocation of Initial Attention

On this attention view, upon inspecting the first stimulus, people are likely to recognize the crucial dimensions, such as size and brightness. (People can identify relevant dimensions, as measured by a feature listing task, before categorization commences; Clapper & Bower, 1994). We suggest that people differentially distribute their attention between those dimensions, sometimes favoring size and sometimes brightness. Although final attention weights can be empirically confirmed (e.g., Kruschke & Blair, 2000), its initial distribution is difficult to measure independently. Instead, hypotheses about initial attention weights can be explored by modeling. The modeling-of-attention approach has accounted for people's categorization strategies in supervised categorization (e.g., Yang & Lewandowsky, 2004), and here we applied a similar technique based on Pothos and Chater's (2002) simplicity principle.

We augmented Pothos and Chater's (2002) simplicity principle with a selective attention mechanism. We varied the amount of attention paid to the binary dimension from .1 (almost none) to .9 (almost exclusive) in steps of .1. Attention to the quinternary dimension concomitantly varied from .9 to .1. For each setting of attention, we computed the compression associated with the binary and quinternary classifications. Compression represents the number of bits required to describe the inequalities among all stimuli given the particular category structure as a proportion of the number of bits required to describe the same inequalities in the absence of any classification (for details, see Pothos & Chater, 2002). The greater the compression, the "simpler" is the category structure and the more it should be preferred. The results (not shown in detail here) were clear: As attention shifted towards the binary dimension, use of a binary classification strategy resulted in increasing simplicity, with a concomitant increase in cost for the quinternary strategy. Conversely, as attention shifted away from the binary dimension, the quinternary strategy yielded greater simplicity at the expense of the binary alternative. It follows that if people commence the task with their attention randomly biased towards one dimension or the other, the

classification strategy associated with that dimension—be it binary or quinternary—would yield

the simpler, and hence preferred, classification.

<center>Unsupervised and Supervised Learning</center>

Our comparison of supervised and unsupervised learning can be summed up in two ways:

First, the rapid unsupervised strategy acquisition observed in both studies stands in contrast to the

rather slow supervised learning in Experiment 1. Not only were supervised participants slower to

learn within the first few trials, but their performance continued to lag behind as far as 20 trials

into the task. The slow supervised learning is quite counter-intuitive because, particularly for the

binary categorizers, it is curious why people would not "click" and catch on to the fact that the

single dichotomous dimension was relevant. In addition, the striking difference in learning rates

between the two modes of categorization could be taken to imply some fundamental processing

differences. The second perspective on our results, however, suggests otherwise. When only those

supervised participants are considered whose first responses was correct, the unsupervised speed

advantage was virtually eliminated. Not only were there no statistical differences between the two

conditions, but the progress of learning with respect to both principal strategic alternatives was

virtually identical for supervised and unsupervised learners (see Figure 3).

What, then, might explain this rather striking pattern of differences and similarities between

the two modes of categorization? Schyns (1991) offered the intriguing suggestion that *all* of

categorization is inherently unsupervised, and that what is commonly thought to be supervised

categorization only extends to the learning of appropriate category names. Actual classification,

by contrast, is achieved by a self-organizing unsupervised system. We adopt this notion to

account for our results: Supervised learning is generally slower than unsupervised learning

because of the added need to associate stimuli with category labels. When the first supervised

trial successfully pairs a label with a stimulus, this initial association provides a reference point

that is identical to what is always available in unsupervised classification. In consequence,

supervised and unsupervised learning rates no longer differ much, if at all. We therefore conclude

that supervised and unsupervised category learning share not only common properties but, quite likely, also some common underlying processes.

This conclusion stands in striking contrast to the interpretations offered by some earlier researchers, who stated that "... category learning was strikingly different, depending on whether feedback was provided… [Unsupervised participants] were only able to learn unidimensional rules…even when such rules failed to separate the stimuli into obvious clusters" (Ashby et al., 1999, p. 1197). Likewise, Ashby et al. concluded that: "When the optimal rule was diagonal, [participants] responded with unidimensional or conjunctive rules ... even under conditions in which optimal performance was quickly achieved when feedback was provided" (p. 1194). In our view, those earlier conclusions were based on a less appropriate conception of unsupervised categorization that examined performance with respect to a single strategy that the experimenter considered correct but kept hidden from participants. When people's unsupervised performance is instead examined with respect to the strategy they clearly prefer a very different pattern emerges: people are able to acquire a multi-dimensional classification strategy (Experiment 1); people retain considerable flexibility and continue to adapt to changes in the stimulus environment (Experiment 2); people's unsupervised strategy acquisition does not appreciably differ from supervised learning under comparable conditions (Experiment 1); and finally, unsupervised strategy acquisition represents a clear instance of true learning, rather than merely prolonged guessing (Experiments 1 and 2).

References

Ashby, F., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in

unsupervised categorization. *Perception & Psychophysics, 61*, 1178-1199.

Billman, D. & Knutson, J. (1996). Unsupervised concept learning and value systematicity: A

complex whole aids learning the parts. *Journal of Experimental Psychology: Learning,*

Memory, & Cognition, 22, 458-475.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433-436.

Clapper, J. P. & Bower, G. H. (1991). Learning and applying category knowledge in unsupervised

domains. *The Psychology of Learning & Motivation, 27*, 65-108.

Clapper, J. P. & Bower, G. H. (1994). Category invention in unsupervised learning. *Journal of*

Experimental Psychology: Learning, Memory, & Cognition, 20, 443-460.

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th Ed.). New York: Harper Collins.

Fried, L. S. & Holyoak, K. J. (1984). Induction of category distributions: A framework for

classification learning. *Journal of Experimental Psychology: Learning, Memory, &*

Cognition, 10, 234-257.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of*

Experimental Psychology: General, 123, 178-200.

Gureckis, T. M. & Love, B. C. (2003). Human unsupervised and supervised learning as a

quantitative distinction. *International Journal of Pattern Recognition and Artificial*

Intelligence, 17, 885-901.

Haubensak, G. (1992). The consistency model: A process model for absolute judgments. *Journal*

of Experimental Psychology: Human Perception & Performance, 18, 303-309.

Imai, S. (1966). Classification of sets of stimuli with difference stimulus characteristics and

numerical properties. *Perception & Psychophysics, 1*, 48-54.

Kalish, M. L., Lewandowsky, S., & Davies, M. (2005). Error-Driven Knowledge Restructuring in

Categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition,*
31, 846-861.

Kruskal, J. B., Young, F. W., & Seery, J. B. (1978). *How to use KYST-2, a very flexible program*
to do multidimensional scaling and unfolding. Murray Hill, NJ: Bell Laboratories.

Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science*, *8*,
225-247.

Kruschke, J. K. & Blair, N. J. (2000). Blocking and backward blocking involve learned

inattention. *Psychonomic Bulletin and Review, 7*, 636-645.

Lewandowsky, S., Kalish, M., & Griffiths, T. L. (2000). Competing strategies in categorization:

Expediency and resistance to knowledge restructuring. *Journal of Experimental*
Psychology: Learning, Memory, & Cognition, 26, 1666-1684.

Love, B. C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic*
Bulletin & Review, 9, 829-835.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A Network Model of Category

Learning. *Psychological Review, 111*, 309-332.

Macho, S. (1997). Effect of relevance shifts in category acquisition: A test of neural networks.

Journal of Experimental Psychology: Learning, Memory, & Cognition,23, 30-53.

Maddox, W., Filoteo, J., Lauritzen, J., Connally, E., and Hejl, K. D. (2005). Discontinuous

categories affect information-integration but not rule-based category learning. *Journal of*
Experimental Psychology: Learning, Memory, & Cognition,31, 654-669.

Milton, F. & Wills, A. (2004). The influence of stimulus properties on category construction.

Journal of Experimental Psychology: Learning, Memory, & Cognition 30, 407-415.

Pelli, D. G. (1991). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437-442.

Pothos, E. M. & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science, 26*, 303-343.

Pothos, E. M. & Chater, N. (2005). Unsupervised categorization and category learning. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 58A*, 733-752.

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

Stewart, N., Brown, G. D., & Chater, N. (2002). Sequence effects in categorization of simple perceptual stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 28*, 3-11.

Triandis, H. & Berry, J. (1980). *Handbook of cross-cultural psychology* (Vol. 2). Boston: Allyn & Bacon.

Yang, L.-X. & Lewandowsky, S. (2004). Knowledge partitioning in categorization: Constraints on exemplar models. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30*, 1045-1064.

Author Note

Footnotes

[1] Three participants provided three or more ambiguous stimulus assignments (i.e., equal number of A and B responses). These people were deemed unsure of their final categorization strategy and were thus removed from further analyses.

[2] The remaining 9 (19%) participants used a variety of other idiosyncratic strategies, with a maximum of two participants using the same strategy, and were thus not considered further. No supervised training of those idiosyncratic strategies was performed.

[3] Although the focus of this analysis is on the unsupervised participants, we have included the supervised data as a comparison, using all participants rather than the first-response-correct group only. The predicted values of both groups are equivalent for binary categorizers, but for the quinternary categorizers, the supervised and unsupervised predicted values differ slightly. This is due to there being a single training strategy for the supervised quinternary participants, whereas for the unsupervised participants a mixture of strategies must be taken into account. The observed differences in predicted values were so slight that for clarity of exposition only a single (aggregate) line is shown in the figure.

[4] We thank an anonymous reviewer for alerting us to the possibility that unsupervised participants were simply guessing rather than learning anything about the category structure.

[5] These individuals were excluded from the gap condition due to three or more ambiguous stimulus assignments.

[6] Two individuals in Phase 1 chose to employ a multi-dimensional strategy; however, as in Experiment 1, they were considered together with the quinternary categorizers.

[7] A subsequent ANOVA including the last few trials of Phase 2 in addition to the earlier trial periods was also carried out; however, this did not change the significance of any effects and is therefore not reported.

[8] This analysis was not performed for the binary dimension because both binary dimensional values occurred equally often on the first trial, which irrespective of their final classification strategy most participants invariably assigned to category "A." Hence the binary analysis would carry no informational value.

Table 1

Numbers (Percentages) of Participants in Each Condition of Experiment 2 Who Utilized Each

Categorization Strategy

	Condition				
	Extra Column	Gap	Balanced	Balanced Gap	Overall
Phase 1					
Size-based (binary)	16 (73)	16 (67)	19 (73)	18 (75)	69 (72)
Brightness-based					
1 │ 2 3 4 5 or 1 2 3 4 │ 5	2 (9)	1 (4)	-	1 (4)	4 (4)
1 2 │ 3 4 5 or 1 2 3 │ 4 5	4 (18)	7 (29)	7 (27)	5 (21)	23 (24)
Total	6 (27)	8 (33)	7 (27)	6 (25)	27 (28)
Phase 2					
Size-based					
1 │ others [a]	7 (32)	8 (33.5)	9 (35)	3 (12.5)	27 (28)
1 2 │ others	5 (22.5)	7 (29)	7 (27)	12 (50)	31 (32)
others │ 5	n/a	n/a	2 (8)	1 (4)	3 (3)
Total	12 (54.5)	15 (62.5)	18 (69)	16 (66.5)	61 (63.5)
Brightness-based					
1 │ 2 3 4 5 or 1 2 3 4 │ 5	1 (4.5)	2 (8.5)	-	-	3 (3)
1 2 │ 3 4 5 or 1 2 3 │ 4 5	9 (41)	7 (29)	8 (31)	8 (33.5)	32 (33)
Total	10 (45.5)	9 (37.5)	8 (31)	8 (33.5)	35 (36.5)

[a] The term "others" refers to all dimensional values on one side of the boundary, whether contiguous or

discontinuous.

Table 2

Proportion of Times Each Quinternary Dimension Value Appears on the First Trial for

Binary and Quinternary Categorizers in Experiment 1 and Phase 1 of Experiment 2

	Quinternary Dimension Value				
	1	2	3	4	5
Experiment 1					
Binary ($N = 17$)	.35	.12	.18	.18	.18
Quinternary ($N = 22$)	.32	.23	.04	.27	.14
Experiment 2: Phase 1					
Binary ($N = 69$)	.26	.16	.22	.17	.19
Quinternary ($N = 27$)	.30	.15	.22	.22	.11

Figure Captions

Figure 1.   Multidimensional scaling (MDS) solution for the complete stimulus matrix used in both experiments based on the similarity ratings by a separate group of subjects who never categorized the stimuli. The dashed and dotted boxes represent, respectively, the size-binary and brightness-binary stimuli for Experiment 1.

Figure 2.   Proportion of responses during trials 1-3, 3-5, 8-10 and 18-20 that matched people's ultimate classification strategy for binary categorizers (top panel) and quinternary categorizers (bottom panel) in Experiment 1. Solid lines with squares represent unsupervised participants and dashed lines with triangles represent all supervised participants.

Figure 3.   Proportion of responses matching participant's chosen strategy plotted against the proportion matching the alternative strategy for binary (top panel) and quinternary (bottom panel) categorizers in Experiment 1. Solid lines with circles represent unsupervised categorizers, and dotted lines with unfilled circles represent maximally comparable supervised categorizers whose first response was correct. Performance at the outset (trials 1-3) is represented by the larger plotting symbols and subsequent learning is represented by the smaller plotting symbols connected by arrows (with the final symbol representing trials 18-20).

Figure 4.   Proportion of times category assignment switched between consecutive trials in response to a difference in quinternary value for binary categorizers (top panel) and quinternary categorizers (bottom panel) in trials 1-5, 8-12 and 16-20. Dotted lines represent the predicted switch proportions based on perfect adherence to final strategy.

Figure 5.   Proportion of responses during trials 1-3, 3-5, 8-10 and 18-20 of Phase 1 in

Experiment 2 that matched people's ultimate classification strategy for Phase 1 (top

panel), and the proportion of responses during the same trials of Phase 2 that matched

participant's initial Phase 1 (middle panel) and Phase 2 (bottom panel) strategies.

Figure 6.   Proportion of "A" responses to the first presentation of each "transfer" stimulus in

Phase 2 of Experiment 2, aggregated over all conditions. The top panel shows

transfer responses for people who were classified as binary categorizers in Phase 1,

the bottom panel for quinternary categorizers.

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6