Memory & Cognition
2006, 34 (7), 1398-1413

Ad hoc category restructuring

DANIEL R. LITTLE and STEPHAN LEWANDOWSKY
University of Western Australia, Crawley, Western Australia

and

EVAN HEIT
University of California, Merced, California

Participants learned to classify seemingly arbitrary words into categories that also corresponded to ad hoc categories (see, e.g., Barsalou, 1983). By adapting experimental mechanisms previously used to study knowledge restructuring in perceptual categorization, we provide a novel account of how experimental and preexperimental knowledge interact. Participants were told of the existence of the ad hoc categories either at the beginning or the end of training. When the ad hoc labels were revealed at the end of training, participants switched from categorization based on experimental learning to categorization based on preexperimental knowledge in some, but not all, circumstances. Important mediators of the extent of that switch were the amount of performance error experienced during prior learning and whether or not prior knowledge was in conflict with experimental learning. We present a computational model of the trade-off between preexperimental knowledge and experimental learning that accounts for the main results.

When people learn new categories, they rely not only on empirical observation of category members, but also on prior expectations and beliefs about the categories. The effects of preexperimental knowledge are varied and range from providing an initial concept representation to informing which features or dimensions are relevant for categorization (see, e.g., Heit, 1997; Murphy & Medin, 1985; Wisniewski, 1995). Preexperimental knowledge may have greatly different effects, depending on when during the course of learning it becomes available. When preexperimental knowledge is available from the outset—for example, in the form of category labels—it guides learning by providing an explanation for the structure and properties of categories (Kaplan & Murphy, 2000). When preexperimental knowledge is not explicitly provided at the outset, people use the category members presented during experimental learning to determine which preexperimental knowledge, if any, is useful (see, e.g., Heit & Bott, 2000; Heit, Briggs, & Bott, 2004).

To date, research has primarily focused on the integrated use of preexperimental and experimental knowledge during learning of new categories (Heit, 1994; Heit & Bott, 2000; Kaplan & Murphy, 2000; Wisniewski, 1995). Although the integrated use of different types of knowledge undoubtedly occurs in many real-world situations, there may be other instances in which one source of knowledge competes with and potentially replaces another source. Consider, for example, learning to play a new game of cards with a friend. Initially, you might concentrate on memorizing your friend's instructions. But over time, you might realize that this new card game is just like some other familiar game, with a few small changes. In this case, you may abandon the memorized instructions and continue playing by relying on your prior knowledge of the familiar game.

We examined the issue of how people might switch between experimental learning and preexperimental knowledge by conducting experiments with ad hoc categories (Barsalou, 1983)—that is, categories such as "things that you take from a house that is on fire" that dynamically link otherwise disparate items, such as pets, passports, and photo albums. Ad hoc categories allow a set of items to be represented as unrelated or related, depending on whether or not the ad hoc label is available. For example, Barsalou showed that when category labels were withheld, free recall of ad hoc category members did not differ from the recall of random word lists (and recall of both was worse than recall of lists drawn from common categories). It follows that the observation of ad hoc category members in the absence of the ad hoc label is unlikely to permit the application of useful preexperimental knowledge. Instead, when the label is withheld, categorization of ad hoc category members should rely largely on experimental knowledge attained during feedback-based training.

By implication, if participants are trained to categorize ad hoc category members before the labels are eventually revealed, we can examine how categorization strategies change as people switch from using experimentally ac-

Preparation of this article was facilitated by a Discovery Grant from the Australian Research Council to the second author and Mike Kalish. We thank Leo Roberts and Ed Wilson for their assistance during data collection and manuscript preparation. Correspondence may be addressed to S. Lewandowsky, School of Psychology, University of Western Australia, Crawley, WA 6009, Australia (e-mail, lewan@psy.uwa.edu.au; Web site, www.psy.uwa.edu.au/lewan/lewan.htm).

quired knowledge to using preexperimental knowledge (or to a blend of both types of knowledge). This potential change in strategy, with no change in the nature of the task or the stimuli, is known as *knowledge restructuring* and has previously been examined in categorization using artificial stimuli (see, e.g., Kalish, Lewandowsky, & Davies, 2005; Lewandowsky, Kalish, & Griffiths, 2000). In the studies by Kalish et al. (2005) and Lewandowsky et al. (2000), restructuring from a simple (but imperfect) to a complex (but potentially perfect) categorization strategy occurred only if two conditions were met: First, the complex alternative strategy had to be explicitly pointed out to participants, and second, participants using the simple strategy had to be committing a considerable number of errors. When one or the other condition was not met, people resisted knowledge restructuring and continued to use the simple but imperfect strategy.

We adapted the knowledge restructuring methodology (Kalish et al., 2005; Lewandowsky et al., 2000) to examine the interplay of experimental and preexperimental knowledge in ad hoc categorization. We compared performance across three conditions: a condition in which the ad hoc category labels were available throughout the experiment, a control condition in which the ad hoc labels were never shown, and a condition in which the ad hoc labels were revealed after training. Emphasis was on this "reveal" condition, in which participants had the option of shifting their categorization strategy from one based on experimental learning to one based on preexperimental knowledge.

To assist in identifying strategy use, one item from each ad hoc category was randomly assigned to a different category during training. These "anomalous" items were diagnostic because experimental and preexperimental knowledge dictated different categorization responses. Accordingly, in the reveal condition, restructuring would be identified if people no longer categorized the anomalous items according to training but instead placed them into their ad hoc categories once the ad hoc labels had been revealed. The extent of restructuring could be assessed by contrasting postreveal performance with that in the control condition, in which anomalous items should only be categorized as trained, and that in the throughout condition, in which anomalous items should be categorized into the appropriate ad hoc category at the outset (although experimental learning might override those responses eventually).

At least three competing accounts provide predictions for restructuring of ad hoc categories. The first account, based on the notion of simplicity (Feldman, 2003; see also Pothos & Chater, 2002), predicts that by revealing the simpler of two options after training, participants will be strongly compelled to restructure their knowledge, irrespective of the level of error. This *simplicity* approach specifies independent and unambiguous criteria of complexity (e.g., criteria based on minimum description length; Pothos & Chater, 2002). Accordingly, a verbatim list of seemingly unrelated category exemplars (e.g., "passport," "jewelry," "pets") would constitute an undesirably complex representation, whereas the presence of an ad hoc label ("things to take out of a burning house") would permit the same exemplars to be described in a highly compressed—and hence simpler—manner. The simplicity account thus predicts a complete abandonment of experimental knowledge in favor of the simpler ad hoc category under all circumstances.

The second account invokes a simple model of associative learning and predicts that people should continue to use experimental knowledge if it has successfully supported classification of experimental stimuli (see, e.g., Hall, 1991). For this associative-learning account, all items should continue to be categorized into the appropriate training category if they are sufficiently learned. Although this basic idea is compatible with the results of Lewandowsky et al. (2000) and Kalish et al. (2005), a simple associative-learning view is constrained to expect identical outcomes for all item types, irrespective of whether or not experimental and preexperimental knowledge are in conflict.

The final account also acknowledges the role of learning and the necessity of error for knowledge restructuring to occur, but it also considers the role of prior knowledge in more detail. We formalize these ideas below in a computational model in which experimental learning has the twofold role of establishing associations between items and responses while also selecting whichever prior knowledge is relevant to the task. In consequence, the model can differentiate between situations in which prior knowledge is in conflict with experimental knowledge (i.e., for the anomalous items) and those in which it is not. The present results turn out to be quantitatively compatible with this model and, by the same token, turn out to be at least partially in conflict with the first two alternatives.

EXPERIMENT 1

Experiment 1 examined knowledge restructuring of ad hoc categories after limited training. We expected the remaining performance error to be sufficient for a strategy shift to be observed once the ad hoc labels were revealed. All conditions began with an initial transfer test, followed by a training phase, and concluded with two further transfer tests (which we call "posttraining" and "final" from here on).

Participants in the control condition were not informed of the ad hoc category labels at any time. Participants in the reveal condition were provided with the mapping between ad hoc categories and the arbitrary labels (i.e., A = "Things you take from a house that is on fire"; see Table 1) after the posttraining transfer test, and the throughout condition was given the ad hoc labels at the outset, before the initial transfer test.

Method

Participants. Forty-five University of Western Australia students received course credit for participation. Assignment to conditions

Table 1
Ad Hoc Category Stimulus Lists

Things You Take From a House That Is on Fire	Things That Can Fall on Your Head	Things You See in a Police Station	Things You Can Buy at a Petrol Station
Stimuli Used in All Experiments			
Camera	Apples	Bars	Cheese
Cats	Birds	Cells	Cigarettes
Children	Bombs	Chairs	Dog Food
Clothes	Brick	Computers	Film
Documents	Confetti	Criminals	Gum
Heirlooms	Dirt	Desks	Ice Cream
Jewelry	Leaves	Donuts	Newspaper
Money	Meteors	Doors	Petrol
Pictures	Rocks	Fingerprints	Map
Records	Snow	Paperwork	Soda
Stereo	Water	Uniforms	Toilet Paper
Additional Stimuli Used in Experiment 1			
Dogs	Hail	Badges	Chips
Family	Missiles	Cops	Chocolate
Memorabilia	Rain	Forms	Coffee
Possessions	Sleet	Handcuffs	Jerry Can
TV	Tree Branch	Mugshots	Tire Gauge

Note—In Experiment 1, “Map” was presented as “Road Map.”

was random, with 15 participants in the control condition, 16 in the reveal condition, and 14 in the throughout condition.

Stimuli and Apparatus. Four 16-item sets were prepared for the following ad hoc categories: “Things you take from a house that is on fire,” “Things that can fall on your head,” “Things you see in a police station,” and “Things you can buy at a petrol station” (see Table 1). Separate one-way ANOVAs for familiarity, concreteness, frequency, and imageability statistics (Wilson, 1988) found no significant differences between the categories, with the largest F ratio being for frequency [$F(3,60) = 2.53$, $p > .05$]. For each participant, 1 randomly chosen item was removed from each category and assigned to a different randomly chosen category. Hence, each set consisted of 1 such anomalous item and 15 true items; the true items are referred to from here on as *old* or *new*, depending on (respectively) whether the item was shown during training and transfer or during transfer only.

Each participant received a training list that comprised the anomalous items plus a random selection of 40 words from the remaining stimuli. The transfer list was composed of all items from each category—that is, 40 old, 4 anomalous, and 20 new items. A computer was used to display stimuli and record responses.

Procedure. On each training trial, a randomly chosen training item was presented that participants had to categorize into one of four categories (by pressing one of four keys labeled *A*, *B*, *C*, or *D*). Responses were followed by feedback (“correct” or “wrong”) for 500 msec. Trials were separated by 500-msec blank intervals.

Training consisted of five blocks of trials, each involving presentation of the 44 training items in a different random order. Transfer tests involved presentation of the 64 test items in random order, with a 500-msec blank interval between each response and the next item. Transfer responses received no feedback. After the posttraining transfer test, the participants in all conditions were again instructed to categorize the items on the final transfer test in accordance with how they were trained. At this point, the participants in the reveal condition were given a diagram of how the ad hoc categories mapped to the arbitrary categories.

Results

Three participants failed to categorize old and anomalous items above chance level (.25) in the posttraining transfer test. These participants were excluded from further analysis, leaving 14 participants in each condition. To focus presentation, we report only the highest-order significant effects for all of the ANOVAs in this article. Significant subordinate component effects (e.g., main effects within a two-way interaction) are provided in footnotes but are not discussed unless there is a compelling reason to do so.

Training performance. Figure 1 shows performance during training. In the control and reveal conditions, clear learning was evident for old items. Anomalous items exhibited the same trend, but their performance lagged considerably behind the old items. In the throughout condition, old items were nearly always categorized as trained from the outset, whereas anomalous items initially were rarely categorized as trained (instead being categorized in accord with preexperimental knowledge). Anomalous items were classified into the trained category only later in learning.

Statistical confirmation of those effects was provided by a 3 (condition) × 5 (block) × 2 (item type) ANOVA. The three-way interaction of all variables [$F(8,156) = 8.33$, $MS_e = 0.02$, $p < .05$, $\eta_p^2 = .30$] reflected the fact that the difference between old and anomalous items increased across training blocks in the control and reveal conditions but decreased in the throughout condition.[1]

Transfer performance. By the final transfer test, more than 80% of all items were classified either into the trained or the ad hoc category, rather than into one of the two alternative categories. Hence, the results were nearly completely captured by the trade-off between only two proportions—trained versus ad hoc. For ease of presentation, all remaining analyses consider only the proportion of items classified into the trained category.

Figure 2 shows performance on all three transfer tests (solid lines represent model predictions to be discussed

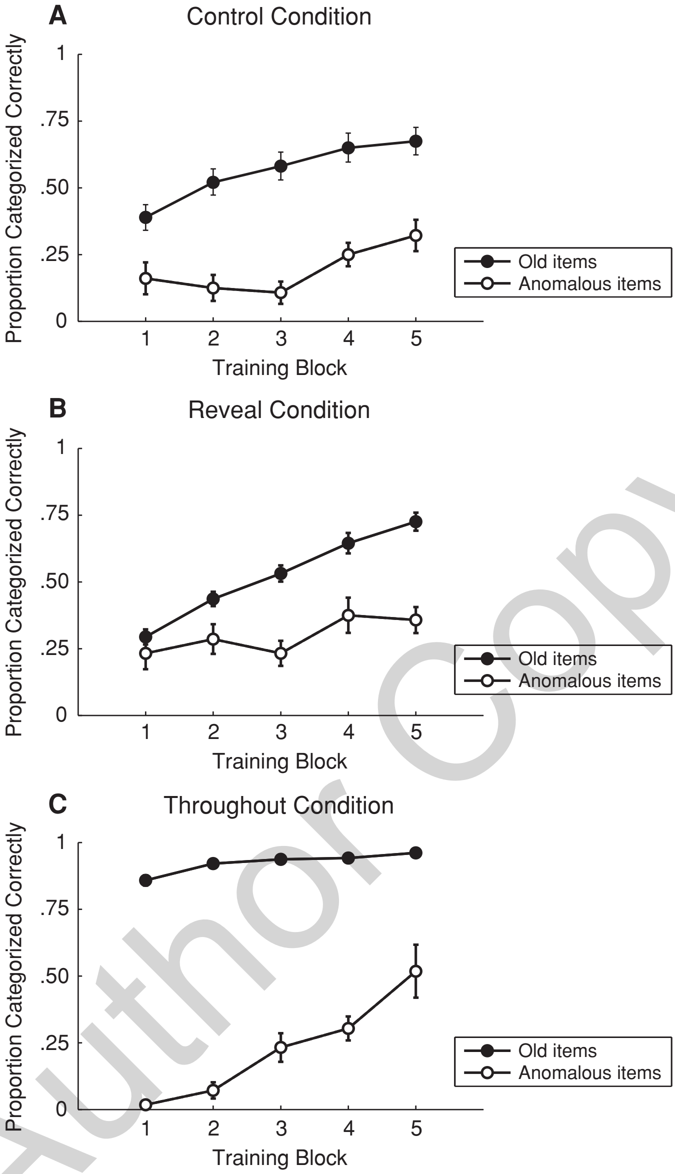

Figure 1. Mean proportions of old and anomalous items categorized correctly in the control, reveal, and throughout conditions during training in Experiment 1.

later). The top panel shows the initial transfer test, before any learning took place. Performance in the control and reveal conditions was at chance (around 25%), whereas people in the throughout condition clearly used preexperimental knowledge, with accurate categorization of old and new items and inaccurate categorization of anomalous items (recall that performance is measured with respect to the *trained* category).

The middle panel shows the posttraining transfer test. In comparison with the initial test, in the control and reveal conditions old items were more likely to be classified in accordance with training. The new items also showed this pattern, suggesting that even without ad hoc labels, preexperimental knowledge exerted some influence during training. In the throughout condition, the main change was the increased tendency to categorize anomalous items to reflect training.

The bottom panel shows the final transfer test, after the ad hoc categories were revealed in the reveal condition. The key results in this condition were increased classification of new items into the trained category and a decreased level of classification of anomalous items in the trained category; both of these results are clear evidence of reliance on preexperimental knowledge.

The classification proportions were analyzed using a 3 (condition) × 3 (test) × 3 (item type) ANOVA. The three-way interaction of all variables was significant, with differences between conditions emerging during the last two transfer tests. The throughout and control conditions showed little change between the posttraining and final transfer tests, whereas the reveal condition showed improved performance on old and new items accompanied by decreased performance on anomalous items, as revealed by the test × condition interaction [$F(4,78) = 2.74$, $MS_e = .02$, $p < .05$, $\eta_p^2 = .12$] and the three-way interaction [$F(8,156) = 27.98$, $MS_e = .02$, $p < .01$, $\eta_p^2 = .59$].[2] These conclusions were confirmed by simple comparisons between the posttraining and final transfer tests in the reveal condition for old [$t(13) = -5.55$, $p < .01$, $r_{Y\lambda} = -.61$], new [$t(13) = -8.47$, $p < .01$, $r_{Y\lambda} = -.78$], and anomalous [$t(13) = 3.48$, $p < .01$, $r_{Y\lambda} = .50$] items.

Discussion

Experiment 1 established the interactive role of preexperimental knowledge and experimental learning. In the throughout condition, preexperimental knowledge dominated, whereas in the control condition, categorization was largely influenced by experimental knowledge, although preexperimental knowledge seemed to play a role in classification of the new items. The latter finding confirms previous reports that participants look for useful preexperimental knowledge even when there are no explicit pointers for doing so (Heit & Bott, 2000; Heit et al., 2004).

The reveal condition produced the most novel results, showing that people's categorization strategies can be readily and dramatically changed under experimental control, when the previously concealed body of preexperimental knowledge is revealed. Before the ad hoc category labels were given, participants categorized largely on the basis of experimental knowledge, but once the labels were given, categorization shifted in the direction of preexperimental knowledge.

EXPERIMENT 2

Training was limited in Experiment 1 to ensure that error remained high, thus facilitating knowledge restructuring. In Experiment 2, the length of training was increased to minimize the remaining performance error. According to the work by Kalish et al. (2005), reveal participants should then have continued to categorize old and anomalous items as trained, as the control group did, despite the revelation of the ad hoc categories. Participants in our throughout condition were only presented with the initial transfer test, for comparison with the reveal condition, but did not engage in any subsequent training.

In addition, the stimuli in Experiment 2 were revised with the intention of reducing the number of meaning-based clusters within each category (i.e., the number of words with similar meanings within each ad hoc category was reduced). This change was expected to attenuate the effects of preexperimental knowledge during uninformed training, thus facilitating detection of knowledge restructuring in response to revelation of the labels.

Method

Forty-five University of Western Australia students were randomly assigned into the reveal, control, and throughout conditions, with equal numbers in each group. The throughout-condition participants completed the initial transfer test only; those in the reveal and control conditions completed all phases of the experiment, which mirrored those of Experiment 1.

To enable more accurate learning than in Experiment 1, the number of items was reduced from 64 to 44 (see Table 1). As in the previous experiment, separate one-way ANOVAs found no differences between the categories for familiarity, concreteness, frequency, and imageability [the largest F ratio was for frequency: $F(3,30) = 1.86$, $p > .05$]. The 24 training items consisted of 4 randomly assigned anomalous items and 20 other items selected at random (5 from each category). The remaining 20 items were shown only during transfer tests.

The procedure was otherwise identical to that of Experiment 1; however, training in the reveal and control conditions was extended either until the participant perfectly categorized all of the training items twice in a row or until 12 training blocks were completed.

Results

One participant in the control condition and 1 in the reveal condition reached the criterion of two consecutive perfect training blocks (after nine and seven blocks, respectively). For those participants, performance in the remaining training blocks was considered to be all correct for the analysis.

Training performance. Figure 3 shows performance during training. In comparison with the findings in Experiment 1, the asymptotic level of performance was higher and the difference between old and anomalous items was smaller. Both features of the results confirm that training here concluded with a lower level of performance error than in Experiment 1.

A 2 (condition) × 12 (block) × 2 (item type) ANOVA nonetheless revealed a significant effect of item type [$F(1,28) = 6.78$, $MS_e = .20$, $p < .05$, $\eta_p^2 = .20$], although its magnitude was considerably smaller than in Experiment 1. There was also a significant effect of block [$F(11,308) = 61.13$, $MS_e = .03$, $p < .01$, $\eta_p^2 = .69$]. No other effects reached significance, with the block × item type interaction yielding the largest F ratio [$F(11,308) = 1.22$].

Transfer performance. Figure 4 shows transfer performance, again represented as the proportion of items classified as trained. There was a strong tendency to follow the response dictated by training when ad hoc category labels were not provided (middle panel). Although providing the ad hoc labels in the reveal condition had some effect, the anomalous items seemed more resistant to a change in classification than they were in Experiment 1: People continued to categorize these items according to training rather than according to preexperimental knowledge.

These data were first analyzed using a 2 (condition: control or reveal) × 3 (item type: old, new, or anomalous) × 3 (transfer test) ANOVA. The significant three-way inter-

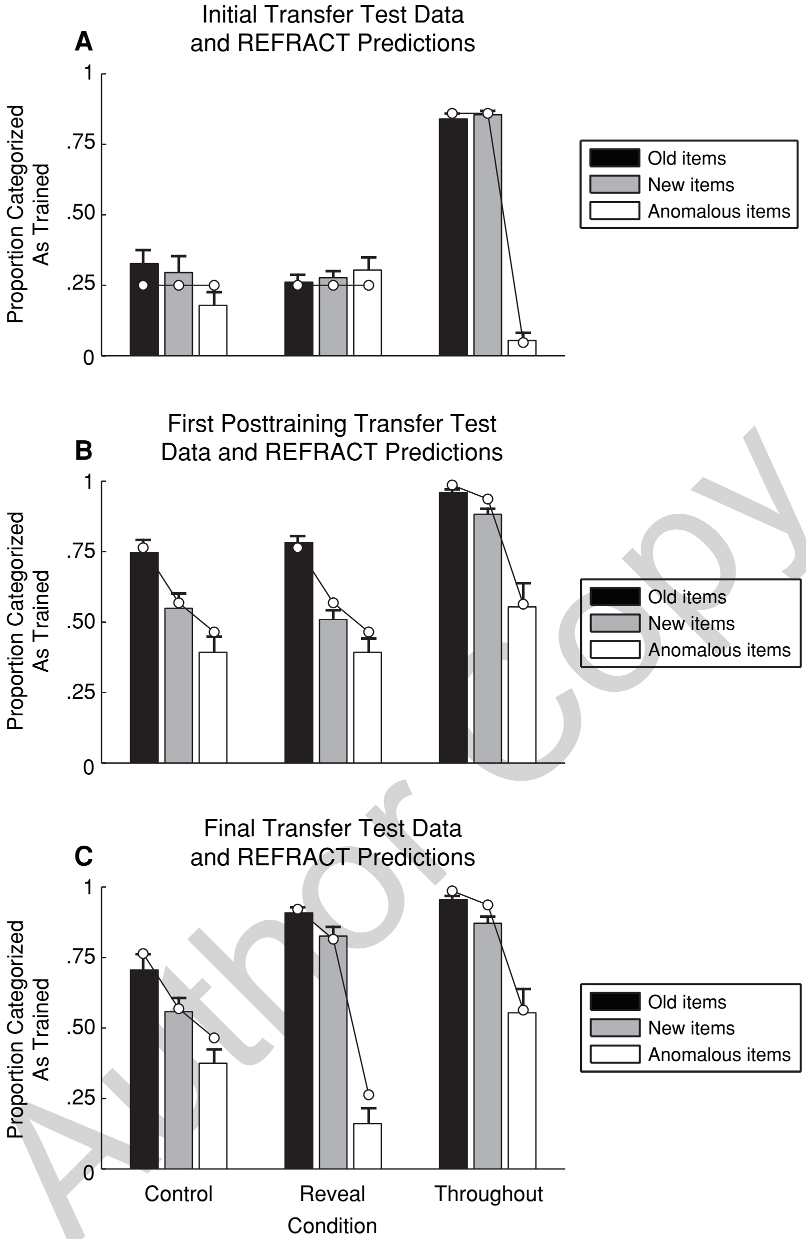

Figure 2. Mean proportions of old, new, and anomalous items categorized as trained in all transfer tests for the control, reveal, and throughout conditions in Experiment 1. The lines represent REFRACT predictions.

action among all variables [$F(4,112) = 18.65$, $MS_e = .02$, $p < .01$, $\eta_p^2 = .40$] reflected differential improvements in performance across transfer tests between conditions.[3]

When the data from the final two transfer tests in the reveal condition were considered in isolation, using a 2 (test) × 3 (item type) ANOVA, there was a significant interaction between test and item type [$F(2,28) = 31.65$, $MS_e = .03$, $p < .01$, $\eta_p^2 = .69$]. This interaction reflected the increase in performance on new items and the accompanying decrease in performance on anomalous items between the posttraining and final transfer tests. This result was further confirmed by simple comparisons of performance on old [$t(14) = -1.07$, $p > .05$], new [$t(14) = -6.73$, $p < .01$, $r_{Y\lambda} = -.73$], and anomalous [$t(14) = 3.24$, $p < .01$, $r_{Y\lambda} = .44$] items between the posttraining and final transfer tests. The corresponding 2 (test) × 3 (item type) interaction failed to reach significance in the control condition [$F(2,28) < 1$]. Hence, we again observed selective knowledge restructuring in the reveal condition, although the decrease in anomalous item performance following

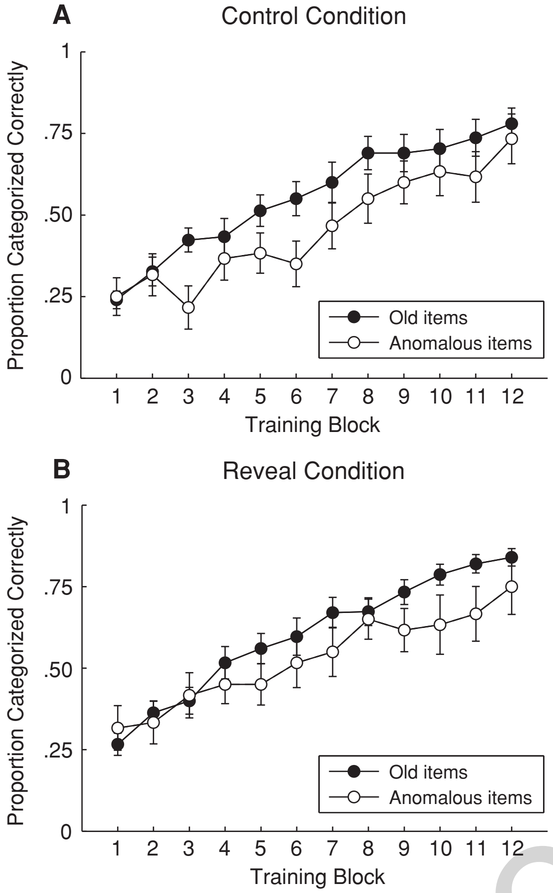

Figure 3. Mean proportions of old and anomalous items categorized correctly in the control and reveal conditions during training in Experiment 2.

the reveal manipulation was not as clear-cut as it had been in the first experiment. This suggests that, in comparison with Experiment 1, the extended training (coupled with a change in stimuli) allowed participants to at least partially resist knowledge restructuring.

Discussion

Together, Experiments 1 and 2 illuminated the way in which experimental learning and preexperimental knowledge competed when both could be used as the basis for responding. In Experiment 1, with its relatively high error after training, participants in the reveal condition readily switched from experimental learning to using preexperimental knowledge after the category labels were revealed. In Experiment 2, with a lower rate of error after training, there was more resistance to restructuring, and participants continued to categorize anomalous items on the basis of experimental knowledge rather than according to their ad hoc status.

However, these conclusions must remain tentative, because Experiments 1 and 2 differed in several ways in addition to varying the extent of training. Specifically, (1) the number of items, (2) the ratio of old to anomalous items, and (3) the ratio of old to new items were all lower in Experiment 2. Experiment 3 was designed to isolate the extent of learning as the variable responsible for the different outcomes of the first two studies.

EXPERIMENT 3

In Experiment 3, we manipulated the extent of training in the reveal condition by replicating the shorter training phase from Experiment 1 with the stimulus set from Experiment 2 in one condition and comparing performance in this short-training condition (with 5 training blocks) to that in an extended-training condition (with 14 training blocks). The control condition was not included here.

Method

Thirty-one participants were randomly assigned to a short-training reveal condition ($n = 17$) or an extended-training reveal condition ($n = 14$). Stimuli were identical to those used in Experiment 2, and the procedure was similar to that of the first two studies. Training in the short-training condition was limited to 5 blocks, as in Experiment 1. Training in the extended-training condition, on the other hand, was lengthened to 14 blocks.

Results

Two participants in the short-training condition and 1 in the extended-training condition failed to categorize old and anomalous items above chance on the posttraining transfer test. These participants were excluded from analysis, leaving 15 in the short-training condition and 13 in the extended-training condition.

Training performance. Figure 5 shows training performance in both conditions. Not surprisingly, final performance was better in the extended-training than in the short-training condition. Also, as in the previous experiments, old items were categorized in accord with training to a greater extent than were anomalous items.

For the short-training condition, a 5 (block) × 2 (item type) ANOVA revealed a significant effect of item type [$F(1,14) = 5.67$, $MS_e = .13$, $p < .05$, $\eta_p^2 = .29$], with better performance on old items than on anomalous items. A significant effect of block [$F(4,56) = 7.92$, $MS_e = .02$, $p < .05$, $\eta_p^2 = .36$], combined with the absence of an item type × block interaction [$F(4,56) = 1.65$, $p > .05$], indicated that performance improved across the training blocks at similar rates for both item types.

Similar results were found in the extended-training condition. A significant main effect of item type [$F(1,12) = 6.07$, $MS_e = .07$, $p < .05$, $\eta_p^2 = .34$] again showed that old items were learned better than anomalous items. The main effect of block [$F(13,156) = 24.44$, $MS_e = .04$, $p < .01$, $\eta_p^2 = .67$] and the nonsignificant interaction between both variables [$F(13,156) < 1$] suggest that performance improved in the same manner for both item types.

Transfer performance. Figure 6 shows transfer performance in both conditions. In the posttraining transfer test, the key result was the greater tendency to categorize anomalous items in accordance with training in the extended-training condition than in the short-training condition. In the final transfer test, after the ad hoc category

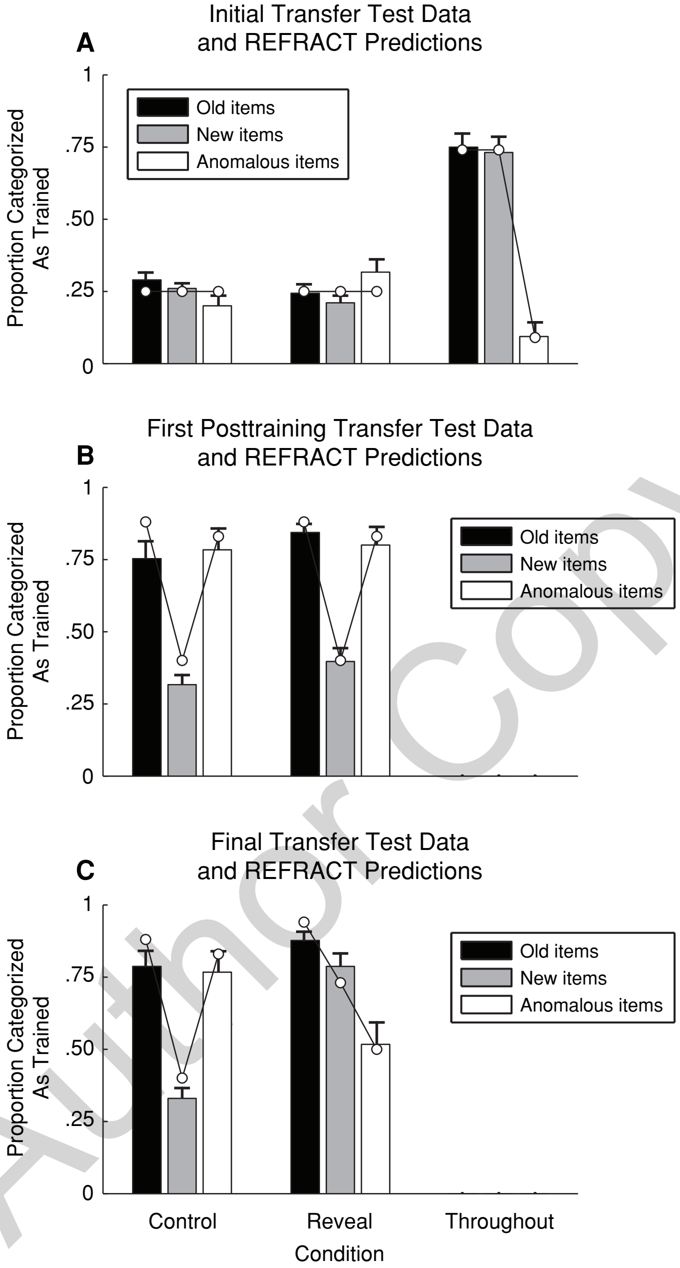

Figure 4. Mean proportions of old, new, and anomalous items categorized as trained in all transfer tests of the control, reveal, and throughout conditions in Experiment 2. The lines represent REFRACT predictions.

labels were revealed, there was extensive restructuring in the short-training condition, such that new and anomalous items were now categorized in accordance with preexperimental knowledge rather than training. In the extended-training condition, by contrast, restructuring was resisted, as evidenced by the continued tendency to classify anomalous items according to training rather than preexperimental knowledge.

A 2 (condition) × 3 (transfer test) × 3 (item type) between- and within-subjects ANOVA revealed a three-way interaction of all variables [$F(4,104) = 8.30$, $MS_e = .02$, $p < .05$, $\eta_p^2 = .24$], which reflected differing degrees of knowledge restructuring in the two conditions.[4] In the short condition, old and new item performance increased in the final transfer test and anomalous item performance decreased to below the level at initial transfer. Simple item

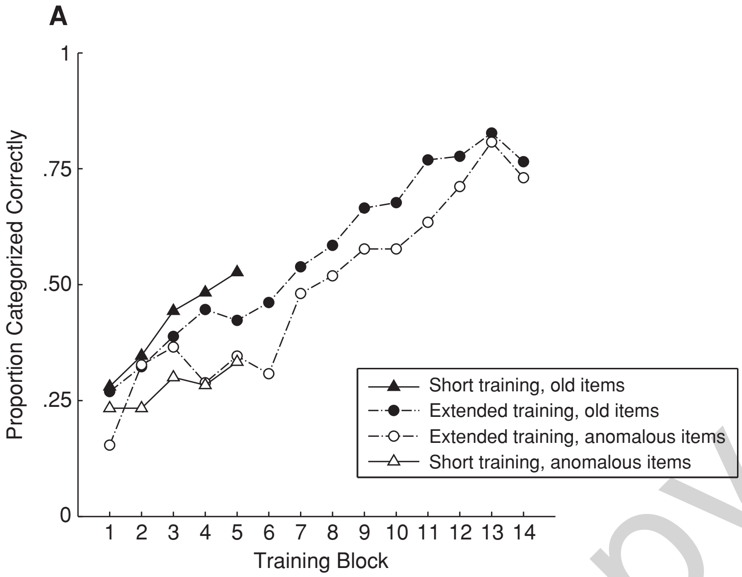

Figure 5. Mean proportions of old and anomalous items categorized as trained in the short and extended reveal conditions during training in Experiment 3. Standard error bars are not shown in order to reduce visual crowding.

comparisons between the posttraining and final transfer tests for the short condition reveal significant differences for old [$t(14) = -5.75, p < .01, r_{Y\lambda} = -.71$], new [$t(14) = -10.54, p < .01, r_{Y\lambda} = -.87$], and anomalous [$t(14) = 4.39, p < .01, r_{Y\lambda} = .61$] items. In the extended condition, old- and new-item performance also increased in the final transfer test, and anomalous-item performance decreased, but only to just below the level observed immediately after training. Here, simple item comparisons between the posttraining and final transfer tests for the extended condition also demonstrated significant differences for old [$t(12) = -1.20, p > .05$], new [$t(12) = -6.38, p < .01, r_{Y\lambda} = -.71$], and anomalous [$t(12) = 3.81, p < .01, r_{Y\lambda} = .32$] items.

Planned comparisons involving only the final transfer test in both conditions (using a 2 × 3 ANOVA) supported these conclusions: In this comparison, there was a main effect of condition [$F(1,26) = 15.13$, $MS_e = .04$, $p < .01$, $\eta^2_p = .37$] and a significant condition × item interaction [$F(2,52) = 16.99$, $MS_e = .04$, $p < .01$, $\eta^2_p = .40$]. This supports the observation that extent of training modulated the effects of revealing the ad hoc labels. One-way ANOVAs across conditions for each item type show that this difference was confined to anomalous items [$F(1,26) = 24.88$, $MS_e = .07$, $p < .01$, $\eta^2_p = .49$]. Performance on old and new items [both Fs$(1,26) < 1$] in the final transfer test was identical across the two conditions.

Correlational analyses. We further explored the relationship between learning and final transfer performance by considering the pattern of correlations between our various measures. For old items, mean performance in the last two transfer tests was identical in the control and reveal conditions across our experiments. However, it is nonetheless possible that the strategy by which old items were classified differed between the two conditions, and this possibility can be examined by looking at the pattern of correlations: If the knowledge used during learning is the same as that used during the final transfer test, then learning and transfer performance should be correlated across participants. By contrast, if restructuring occurred and people shifted to a different strategy after revelation of the ad hoc labels, then the correlation between training and transfer performance should be lower or absent, irrespective of the absolute level of performance.

To maximize statistical power, we combined the data from all three experiments and examined the correlations across participants within the control (Experiments 1 and 2 only) and reveal (Experiments 1–3) conditions separately. An "old-item learning" measure was computed by subtracting initial transfer performance from the posttraining transfer performance for old items. Correlations were then calculated between old-item learning and final transfer performance for old, new, and anomalous items for the reveal conditions and control conditions separately.

Table 2 shows the correlation matrix for both pooled conditions. In the control condition, old-item learning was highly correlated with old-item final transfer. Thus, in the control condition, old-item performance during the final transfer test could be predicted by the degree of learning of those items, which suggests that participants continued to apply experimentally acquired knowledge on the final transfer test. In support, the correlation between old-item transfer and anomalous transfer was also high and marginally significant ($p = .056$) in this condition.

The pattern of correlations in the reveal conditions differed from the control results in one crucial respect:

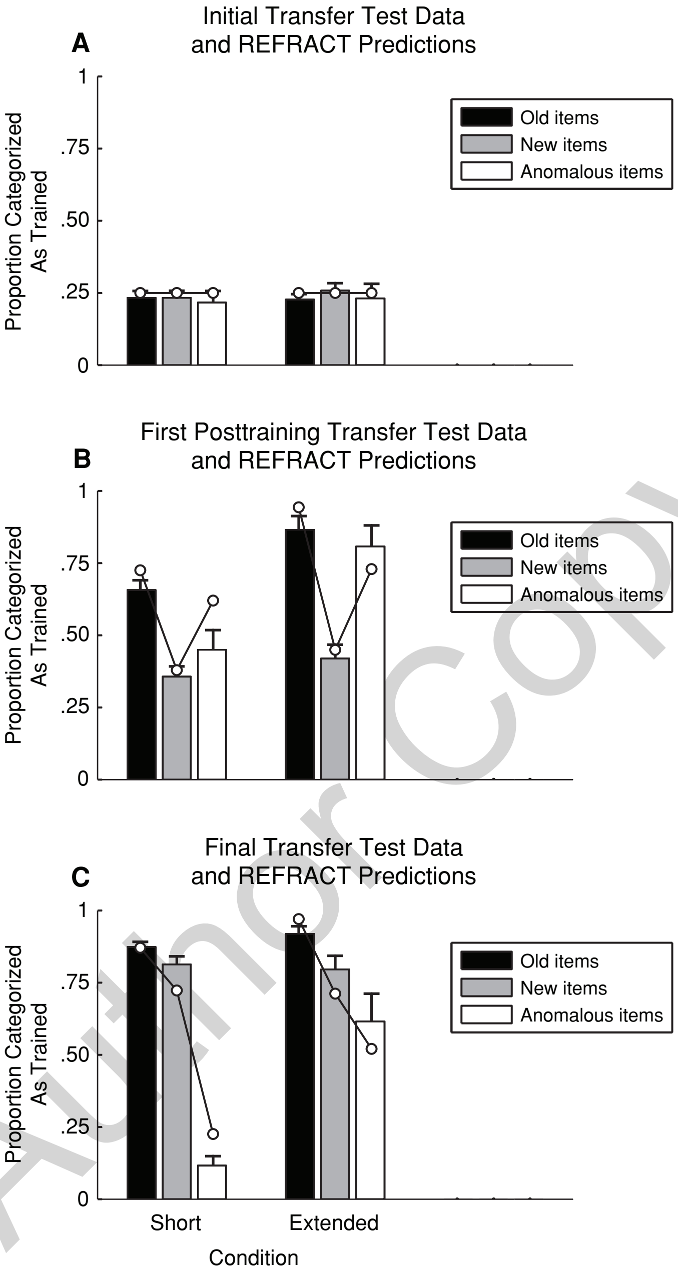

Figure 6. Mean proportions of old, new, and anomalous items categorized as trained in the short and extended reveal conditions during all transfer tests in Experiment 3. The lines represent REFRACT predictions.

Old-item learning was *not* correlated with old-item final transfer performance. Moreover, this correlation was significantly different from the corresponding correlation in the control condition ($z = 2.68$, $p < .01$). This key finding indicates that once ad hoc labels were revealed, participants abandoned experimental knowledge and used preexperimental knowledge to categorize old items. The similar mean performance level on old items in the final transfer test in the reveal and control conditions masked what were actually two different strategies.

The reveal condition thus demonstrated that people relied on different knowledge types for different types of item: For old items, the lack of correlation between learning and final transfer indicates that people shifted strategies when the ad hoc labels were revealed. Anomalous items, by contrast, if well learned during training,

Table 2
Correlations for Old Learning and Final Transfer Test Performance on All Items in the Pooled Control Conditions ($n = 28$) and Pooled Reveal Conditions ($n = 55$) From All Experiments

	1	2	3	4
Control Condition				
1. Old-item learning	–	.72**	.21	.37ⁿ
2. Old-item transfer		–	.29	.52*
3. New-item transfer			–	−.25
4. Anomalous-item transfer				–
Reveal Condition				
1. Old-item learning	–	.25	.08	.53**
2. Old-item transfer		–	.18	.31*
3. New-item transfer			–	−.20
4. Anomalous-item transfer				–

Note—For this analysis, 2 participants were excluded from each (pooled) condition because their influence on the regression was excessive (their Cook's *D* was more than 3 *SD*s from the mean Cook's *D*). These participants were not excluded from the other analyses because their means were within 2 *SD*s of the total cell means. $^*p < .05$. $^{**}p < .01$. $^{n}p = .56$.

were categorized according to training. Taken together, the anomalous item transfer data and the correlations for old items demonstrate selective resistance to knowledge restructuring for anomalous items only; old items, by contrast, were classified on the basis of preexperimental knowledge on the final transfer test.

Discussion

Experiment 3 demonstrated that extended training rendered the experimentally acquired strategy for anomalous items resistant to change. However, that resistance was limited to the anomalous items only; old and new items were classified on the basis of preexperimental knowledge, regardless of the extent of training, once the ad hoc labels were revealed.

Pooling together the results from all experiments, the correlations between training and transfer performance were highly informative. In the control condition, which never received the category labels, the initial degree of learning of the old items was a good predictor of the subsequent tendency on transfer tests to categorize both old and anomalous items according to training. In contrast, in the reveal condition, for old items the corresponding correlation was significantly smaller and not statistically different from zero, indicating that performance was now controlled by preexperimental knowledge. At the same time, learning continued to predict transfer of anomalous items, notwithstanding revelation of the ad hoc labels. This selective use of different knowledge types for different item types during restructuring represents one of the key contributions of our study.

REFRACT: MODELING KNOWLEDGE RESTRUCTURING

Although computational models are central to much categorization research, modeling of preexperimental-knowledge effects has remained relatively scarce. Heit and Bott (2000) proposed a connectionist model, known as Baywatch, that considers preexperimental knowledge in categorization. The model's name refers to its two modes of learning: a knowledge-driven side that follows Bayesian principles, and an experimentally driven side that learns from observations (see Rehder & Murphy, 2003, for a related model). The key contribution of Baywatch is that it uses experimental learning to select prior knowledge. For example, if observed category members resemble office buildings, then prior knowledge about office buildings will be used rather than prior knowledge about other types of buildings.

Because Baywatch is implemented as a mixture-of-experts network (see, e.g., Jacobs, 1997; Kalish, Lewandowsky, & Kruschke, 2004), it seems well suited to capturing trade-offs between preexperimental knowledge and experimental learning. However, Baywatch was only designed for stimuli presented as feature lists, and it cannot model the ad hoc category members in the present experiments. Equally, to our knowledge, no other model has been applied to ad hoc categories. We have therefore developed a new connectionist model, known as REFRACT, that was built on some of the key properties of Baywatch (see the Appendix for implementation details).

Overview of REFRACT

REFRACT represents experimental and preexperimental knowledge in two types of weights that link items to outputs. As shown in Figure 7, input units are connected to output units directly by a fully interconnected set of weights, and also indirectly through a set of preexperimental-knowledge nodes. The direct weights represent experimental learning and accordingly are set to zero at the outset. For the indirect connections, one set of weights, connecting items to preexperimental-knowledge nodes, are pretrained to capture the preexperimental knowledge that is evoked by presentation of an item. The remaining set of weights, between the preexperimental-knowledge nodes and the output layer, represent associations between preexperimental knowledge and particular output categories. The role of these remaining weights is twofold: First, they implement the reveal manipulation by taking on values that capture the relationship between ad hoc categories and preexperimental knowledge (e.g., the fact that "valuables" tend to belong to "things one would take out of a burning house"). Second, they are adjusted during learning when no a priori information about the nature of the output categories is available, which implements the known gradual recruitment of preexperimental knowledge in categorization tasks (see, e.g., Heit et al., 2004).

In REFRACT, the preexperimental-knowledge nodes capture whatever prior knowledge is activated by the items. We have made two assumptions about how items activate prior knowledge that capture the heterogeneous nature of ad hoc categories: First, some items within an ad hoc category activate the same preexperimental knowledge. This consideration was operationalized by allowing

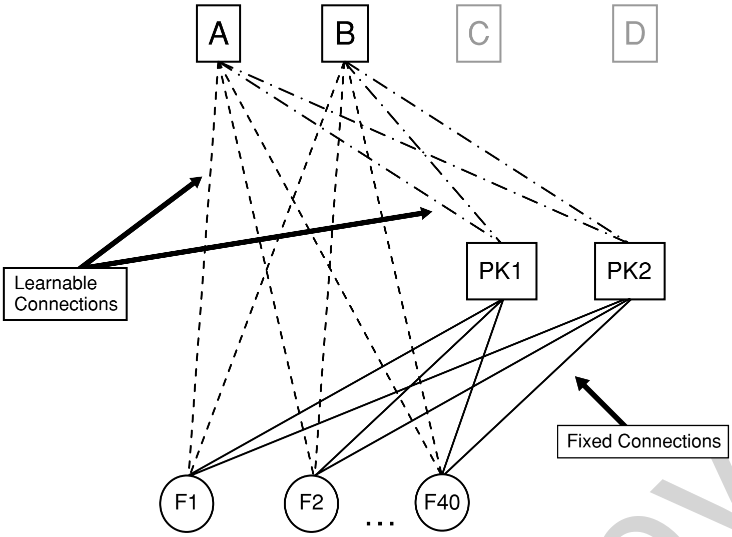

Figure 7. REFRACT model architecture. The circle "F" nodes represent item input features. Preexperimental-knowledge nodes are represented by "PK," and the output categories are denoted by the arbitrary category labels "A," "B," "C," and "D." Note that two preexperimental-knowledge nodes are assigned to each category; in this figure, the nodes of only one category are shown.

items to activate one of two possible preexperimental-knowledge nodes per ad hoc category, thus necessarily enabling multiple items to activate the same prior knowledge. Second, during debriefing our participants indicated that some items did not fit any perceived subcluster, a factor that formed the basis of our second assumption—namely, that some items are highly idiosyncratic and do not seem to share preexperimental knowledge with other items. This second consideration was operationalized by allowing an item to activate none of the preexperimental-knowledge nodes. All preexperimental-knowledge activation is probabilistic and controlled by parameters (see the Appendix for details).

All learnable weights in the model were adjusted using conventional error-driven learning. Output activations were transformed into classification probabilities using Luce's choice rule (see, e.g., Kruschke, 1992).

Model operation. Learning adjusts the weights between items and output nodes and between activated preexperimental knowledge and the output (see Figure 7). For old items, both sets of weights become associated with the same output category. New items are only indirectly associated with the output categories via preexperimental knowledge. Accordingly, when a new item is presented, performance would exceed chance only to the extent that the item activates preexperimental-knowledge nodes that have already been associated with the correct category. Anomalous items require a different response than the category suggested by their activated preexperimental knowledge. Thus, although anomalous and old items both activate the same preexperimental-knowledge node(s), the target output is different between the two item types.

When the ad hoc labels are revealed, the weights between preexperimental-knowledge and all output nodes are incremented by a parameter-determined amount. In consequence, for anomalous items, the reveal manipulation places the learned direct item–output associations in competition with preexperimental knowledge. Knowledge restructuring then occurs if the direct item–output associations are weaker than the competing associations involving preexperimental knowledge. For old items, the item–output connections are congruent with the contribution of preexperimental knowledge, so revelation of ad hoc labels facilitates correct responding, irrespective of the extent of training. New items are facilitated for the same reason.

Application of the model. In simulating Experiment 1, the model identified two effects of preexperimental knowledge during training: First, it caused a deficit for anomalous items (control and reveal conditions) by overshadowing the conflicting experimental learning. Second, it compensated for the absence of learned connections for new items and enabled them to be classified better than anomalous items (see Figure 2, middle panel).

In the simulation of Experiment 2, the decreased role of preexperimental knowledge was modeled by different items from the same ad hoc category activating multiple and idiosyncratic preexperimental-knowledge nodes. Accordingly, performance for new items in the posttraining transfer test (see Figure 4, middle panel) was worse than for anomalous items. Moreover, following the reveal manipulation, performance on anomalous items was correctly predicted to decrease to a lesser extent than in Experiment 1.

The conclusion that knowledge restructuring is a function of the strength of the direct, experimentally learned connections between input and output nodes was confirmed by the simulation of Experiment 3. In both conditions, following the reveal manipulation, simulated performance on old and new items increased to roughly the same level (see Figure 6). By contrast, the drop in anomalous-item performance varied between conditions: In the short condition, the adjustment of the preexperimental-knowledge connections overrode the limited strength of the direct input–output connections, and knowledge restructuring occurred. In the extended condition, the direct input–output connections were sufficiently strong to prevent complete reliance on preexperimental knowledge.

Because the model is constrained to use the same preexperimental-knowledge parameter generated by the simulation of Experiment 2, REFRACT mispredicted anomalous-item performance after training in the short-training condition of Experiment 3; actual performance was lower than predicted. This indicates that the assumption that items activate similar prior knowledge does not capture all aspects of ad hoc learning.

To simulate the individual differences in Experiments 2 and 3, and to confirm the conclusions suggested by the correlations in Table 2, the data from each condition were divided into clusters based on low (.00–.25), midlevel (.50), and high (.75–1.00) final-transfer anomalous-item performance. Parameters were estimated separately for each of these clusters within each condition, and each set of estimates was used to generate 200 random replications. Replications were then aggregated across clusters to provide the full range of simulated individual differences. The pattern of correlations generated by REFRACT matches the empirical pattern (see Table 2). Crucially, the predicted correlation between old-item learning and old-item transfer was considerably higher in the combined control condition ($r = 1.00$) than in the reveal condition ($r = .64$).

GENERAL DISCUSSION

The present results are readily summarized: (1) When useful preexperimental knowledge is revealed, people use it to classify old and new items, regardless of performance on those items, and (2) by contrast, the effects of revelation on anomalous items depends on the extent to which they have been learned. If anomalous items are poorly learned, knowledge restructuring occurs, whereas if anomalous items are well learned, knowledge restructuring is resisted. Crucially, knowledge restructuring is resisted for those items only, implying that experimental and prior knowledge are used simultaneously and selectively for different item types. In addition, the present study shows that providing preexperimental knowledge at the outset enables better learning of exceptions (Experiment 1, throughout condition).

Connections to Previous Research

Experimental and artificial intelligence approaches to preexperimental knowledge. The use of preexperimental knowledge is often seen as a way of facilitating experimental training of categories (Heit & Bott, 2000). However, there are also instances in which preexperimental knowledge and experimental training are in competition. For example, Murphy and Allopenna (1994) and Spalding and Murphy (1999) reported that sensitivity to observed feature frequencies was reduced or eliminated when category members were closely related to preexperimental knowledge.

The present study likewise examined a competitive relation between preexperimental knowledge and experimental learning, although the effects of competition here were found to be asymmetrical. When preexperimental knowledge was available at the start of learning, people were able to update initial expectations in an incremental manner when confronted with conflicting observations. Accordingly, people in the throughout condition gradually acquired the ability to classify the anomalous items into a category other than the one dictated by the ad hoc label.

In contrast, when preexperimental knowledge was revealed after training, people tended to abruptly change the basis of responding, instantly abandoning what they had learned in the experiment (notwithstanding experimental instructions to continue to respond on the basis of training). The only exception to this restructuring involved learned responses that were in conflict with prior knowledge, and even then those learned responses were retained only when they had been learned particularly well.

Some artificial intelligence research (e.g., Geman, Bienenstock, & Doursat, 1992) has emphasized the trade-off between preexperimental knowledge and experimental learning in terms of the bias–variance dilemma. *Bias* refers to imposing constraints from preexperimental knowledge onto what is learned, and *variance* refers to sensitivity to the details and idiosyncrasies of what is learned. Geman et al. argued that as bias increases, variance inevitably decreases, and vice versa. We suggest that the knowledge restructuring observed here is a particularly vivid example of the bias–variance dilemma, in which a sudden increment in the availability of preexperimental knowledge—that is, an increase in bias—entails an equally sudden abandonment of previous experimental learning—a reduction in variance.

Knowledge restructuring. The present research corresponds with some, but not all, previous results on knowledge restructuring in categorization. Turning first to points of agreement, our results confirm the necessity of providing a hint about an alternative strategy, since no restructuring was observed when the reveal manipulation was withheld. This accords well with the findings by Kalish et al. (2005) and Lewandowsky et al. (2000), who similarly found that even after extended learning people did not spontaneously abandon one strategy in favor of another. Moreover, in Lewandowsky et al.'s study, when the alternative strategy was revealed after training, participants resisted knowledge restructuring and continued to use the originally learned strategy. In an exploration of this result, Kalish et al. (2005) showed that resistance to restructuring was directly tied to the amount of remaining performance error at the end of training; in particular, if people were able to learn and memorize exceptional items to a strategy, they persisted with that initial strategy, despite revelation of an alternative. Exactly the same behavior was observed in the present experiments, with people retaining the initial categorization strategy for anomalous items only if they had been given extensive opportunity to learn them.

The present study also diverges from previous knowledge-restructuring research in two ways. First, Lewandowsky et al. (2000) trained their participants to use a simpler, inefficient strategy and then revealed a more efficient but complex strategy. The preexperimental-knowledge strategy in the present experiments, on the other hand, was clearly less complex than the trained strategy. Furthermore, and again unlike Lewandowsky et al.'s design, the revealed ad hoc strategy was also less optimal than continued use of experimental knowledge, in that categorizing anomalous items according to ad hoc labels was erroneous (though, given that there were only four anomalous items, the differences between strategies were small).

Second, after the ad hoc labels were revealed in the present study, participants were found to rely on preexperimental knowledge for old and new items, irrespective of training performance, and error-gated knowledge restructuring was confined to the anomalous items. This result contrasts with Kalish et al.'s (2005) finding that when it occurred, knowledge restructuring was always complete. We suggest that the role of error was so selective here because the learned responses to anomalous items were in conflict with preexperimental knowledge. Thus, when anomalous items were learned well, participants were able to override their prior knowledge and resist knowledge restructuring.

Other uses of ad hoc categories. Although our experiments concentrated on ad hoc category learning in a classification task, there may be aspects of ad hoc categories that are not captured through classification. Different methods of category learning and category use, such as category inference (see, e.g., Yamauchi & Markman, 1998) or problem solving (see, e.g., Ross, 1999), may lead to different representations of ad hoc categories than those explored here.

It remains to be seen whether other forms of representation affect the interaction of experimental and preexperimental knowledge that we observed here.

Potential Alternative Explanations

The present experiments are related to other rule-plus-exception tasks in categorization, except that the rule here was based on preexperimental knowledge. By implication, a model designed to handle rule-plus-exception learning, such as RULEX (Nosofsky, Palmeri, & McKinley, 1994), might be a candidate for explaining our data. Particularly relevant here is RULEX's prediction that exception items should be remembered better than rule-consistent items (Palmeri & Nosofsky, 1995). In Experiment 1, we asked participants at the end of the session to recall all learned items using the ad hoc labels as cues. Because the cued-recall results parallel the transfer results, they were not reported earlier; however, it is noteworthy that in the throughout condition, in which anomalous items were known to be exceptions from the outset, those items (M = .82) were recalled much better than old items (M = .59), as expected by RULEX. Nonetheless, we believe that RULEX is insufficient to deal with all aspects of the data reported here because (1) RULEX is not equipped to handle prior knowledge, and (2) it is constrained to use features of perceptual stimuli as input rather than lists of words.

We now briefly reexamine the simplicity and associative-learning accounts considered at the outset. The simplicity principle predicts that once the ad hoc labels are revealed, people will always use the compressed (hence, simpler) prior-knowledge strategy. This prediction was confirmed in conditions in which training was limited. However, unlike REFRACT, the simplicity principle cannot account for why participants, after additional learning, were willing to maintain the extra complexity of the anomalous items in their postreveal category representations. REFRACT thus provides a better account of the data than does the simplicity principle by specifying the circumstances in which people retain additional complexity.

The second potential alternative explanation would subsume the results under a simple associative-learning umbrella—namely, that performance continues as trained if responses have been learned sufficiently well (see, e.g., Hall, 1991). Indeed, on that basis, the notion that the extent of learning affects knowledge restructuring may appear simplistic at first glance; however, this impression is misleading, because we observed old-item learning to be uncorrelated with performance after revelation of the ad hoc label. The lack of correlation implies that old-item learning is discarded upon revelation of useful preexperimental knowledge, even if the two sources of knowledge mandate identical responses. Only exceptional knowledge is retained, and then only if it has been learned particularly well. No simple model of associative learning would predict such opposing effects of revelation on two classes of items that differ little before the relevance of preexperimental knowledge has become apparent.

REFERENCES

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, **11**, 211-227.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, **127**, 107-140.

Feldman, J. (2003). The simplicity principle in human concept learning. *Current Directions in Psychological Science*, **12**, 227-232.

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, **4**, 1-58.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, **117**, 227-247.

Hall, G. (1991). *Perceptual and associative learning*. Oxford: Oxford University Press, Clarendon Press.

Heit, E. (1994). Models of the effects of prior knowledge on category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 1264-1282.

Heit, E. (1997). Knowledge and concept learning. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts, and categories* (pp. 7-41). London: Psychology Press.

Heit, E., & Bott, L. (2000). Knowledge selection in category learning. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 39, pp. 163-199). San Diego: Academic Press.

Heit, E., Briggs, J., & Bott, L. (2004). Modeling the effects of prior knowledge on learning incongruent features of category members. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 1065-1081.

Jacobs, R. A. (1997). Nature, nurture, and the development of functional specializations: A computational approach. *Psychonomic Bulletin & Review*, **4**, 299-309.

Kalish, M. L., Lewandowsky, S., & Davies, M. (2005). Error-driven knowledge restructuring in categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **31**, 846-861.

Kalish, M. L., Lewandowsky, S., & Kruschke, J. K. (2004). Population of linear experts: Knowledge partitioning and function learning. *Psychological Review*, **111**, 1072-1099.

Kaplan, A. S., & Murphy, G. L. (2000). Category learning with minimal prior knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 829-846.

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

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.

Murphy, G. L., & Allopenna, P. D. (1994). The locus of knowledge effects in concept learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 904-919.

Murphy, G. L., & Medin, D. L. (1985). The role theories in conceptual coherence. *Psychological Review*, **92**, 289-316.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, **101**, 53-79.

Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 548-568.

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

Rehder, B., & Murphy, G. L. (2003). A knowledge-resonance (KRES) model of category learning. *Psychonomic Bulletin & Review*, **10**, 759-784.

Ross, B. H. (1999). Postclassification category use: The effects of learning to use categories after learning to classify. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 743-757.

Spalding, T. L., & Murphy, G. L. (1999). What is learned in knowledge-related categories? Evidence from typicality and feature frequency judgments. *Memory & Cognition*, **27**, 856-867.

Wilson, M. [D.] (1988). MRC Psycholinguistic Database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, **20**, 6-10.

Wisniewski, E. J. (1995). Prior knowledge and functionally relevant features in concept learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 449-468.

Yamauchi, T., & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory & Language*, **39**, 124-148.

NOTES

1. The main effects of condition [$F(2,39) = 19.62$, $MS_e = .08$, $p < .01$, $\eta_p^2 = .50$], item type [$F(1,39) = 219.57$, $MS_e = .09$, $p < .01$, $\eta_p^2 = .85$], and block [$F(4,156) = 49.59$, $MS_e = .02$, $p < .01$, $\eta_p^2 = .56$], as well as the block × item type interaction [$F(4,156) = 2.83$, $MS_e = .02$, $p < .05$, $\eta_p^2 = .07$], were all significant. The block × condition interaction failed to reach significance [$F(8,156) = .93$, $p > .05$].

2. The main effects of condition [$F(2,39) = 39.71$, $MS_e = .07$, $p < .01$, $\eta_p^2 = .67$] and item type [$F(2,78) = 123.52$, $MS_e = .04$, $p < .01$, $\eta_p^2 = .76$], the condition × item type interaction [$F(4,78) = 7.78$, $MS_e = .04$, $p < .01$, $\eta_p^2 = .29$], and the test × item type interaction [$F(4,156) = 11.89$, $MS_e = .02$, $p < .01$, $\eta_p^2 = .23$] were all significant.

3. The main effects of test [$F(2,56) = 124.96$, $MS_e = .04$, $p < .01$, $\eta_p^2 = .82$] and item type [$F(2,56) = 40.14$, $MS_e = .04$, $p < .01$, $\eta_p^2 = .59$] and the interactions item type × condition [$F(2,56) = 6.20$, $MS_e = .04$, $p < .05$, $\eta_p^2 = .18$] and test × item type [$F(4,112) = 24.42$, $MS_e = .02$, $p < .01$, $\eta_p^2 = .47$] were all significant.

4. The main effects of condition [$F(1,26) = 21.52$, $MS_e = .01$, $p < .05$, $\eta_p^2 = .45$], transfer test [$F(2,52) = 183.90$, $MS_e = .03$, $p < .01$, $\eta_p^2 = .45$], and item type [$F(2,52) = 24.13$, $MS_e = .04$, $p < .01$, $\eta_p^2 = .48$] and the interactions condition × transfer test [$F(2,52) = 9.21$, $MS_e = .03$, $p < .01$, $\eta_p^2 = .26$], condition × item type [$F(2,52) = 9.77$, $MS_e = .04$, $p < .01$, $\eta_p^2 = .27$], and item type × transfer test [$F(4,104) = 50.06$, $MS_e = .02$, $p < .01$, $\eta_p^2 = .66$] were also significant.

APPENDIX
Technical Details of REFRACT

Model Implementation

Items were represented by vectors of 40 random features to ensure that all items were orthogonal to each other, each drawn from a uniform distribution with a range [−.5, .5]. When an item was presented, the pattern of activations across the 40 input units was passed along the weights to the output layer, such that activation of each output unit j was the combined total of the weighted sum of all input activations a_i plus the weighted sum of the activations of all preexperimental-knowledge nodes, a_k:

$$O_j = \sum_i \omega_{ij} a_i + \sum_k \omega_{kj} a_k, \tag{1}$$

where ω_{ij} is the association weight between input node i and output node j and ω_{kj} is the association weight between the preexperimental-knowledge node k and output node j. Output activations were mapped into probabilities of responding with category J according to Luce's choice rule

$$\Pr(J) = \frac{\exp\left(\Phi O_J\right)}{\sum_j \exp\left(\Phi O_j\right)}, \tag{2}$$

where Φ is a scaling parameter. A low value of Φ diminishes differences in activation, whereas a high value of Φ accentuates differences in activation (Erickson & Kruschke, 1998).

Input units were connected to preexperimental-knowledge nodes (two for each category) by fixed weights (see Figure 7) that captured the probability with which preexperimental knowledge was activated. The weights were modeled by a single free parameter, $P(\mathrm{A_{pk}})$, that controlled the proportion of items within a category that

APPENDIX (Continued)

activated one of the two preexperimental-knowledge nodes. Specifically, $1 - P(\mathrm{A}_{\mathrm{pk}})$ represented the probability that neither of the preexperimental-knowledge nodes was activated, with each preexperimental-knowledge node being active with probability $P(\mathrm{A}_{\mathrm{pk}})/2$. An item that activated neither of the preexperimental-knowledge nodes assigned to its category was treated as activating an idiosyncratic preexperimental-knowledge node. Preexperimental-knowledge nodes took on activation values of $\{0, 1\}$, representing inactive and active, respectively. Adjustment of $P(\mathrm{A}_{\mathrm{pk}})$ captured the stimulus differences between Experiments 1 and 2.

All other weights were learned via the standard delta rule (see, e.g., Gluck & Bower, 1988)

$$\Delta\omega_{ij} = \beta\left(d_{ij} - O_j\right)a_i, \tag{3}$$

where d_{ij} is the target output for output node j (which is 1 for the correct category and 0 otherwise) given the current input vector, and β is a parameter governing the learning rate.

The reveal manipulation was modeled by the parameter λ, which represented the amount by which the connections between preexperimental-knowledge nodes and output nodes were incremented. In addition, $P(\mathrm{A}_{\mathrm{pk}})$ was set to unity, such that all of the items within a category activated one of the two possible preexperimental-knowledge nodes. This captured the idea that revealing useful preexperimental knowledge would apply to all items equally.

For simulation of the throughout condition, the reveal manipulation was applied at the outset. For the reveal condition, the reveal manipulation was applied after associative weights were adjusted during training and after predictions were generated for the posttraining transfer test. By implication, the reveal and control condition simulations were constrained to make identical predictions prior to the reveal manipulation.

Simulation Regime

Parameters were estimated by minimizing the root-mean squared deviation (RMSD) between predictions and data for training and transfer simultaneously. At each step during the parameter estimation process, the model was run for 200 replications, each with a different item presentation order, but using the same random items to avoid instability during parameter estimation. Table A1 shows the RMSD and parameter values for all simulations. Predictions for the initial transfer test were calculated using those parameter estimates, but after resetting all learnable weights to zero.

Table A1
Model Fits and Parameter Values in REFRACT Simulations

Model Fit	RMSD	β	Φ_{CR}	Φ_{T}	λ	$P(\mathrm{A}_{\mathrm{pk}})$
Experiment 1	.054	.05	1.47	3.22	0.89	1.00
Experiment 2	.047	.03	2.35	1.47	1.46	.63
Experiment 3	.068	.01	4.88		0.83	.63

Using the parameters generated above, with separate Φ values for the control and reveal and for the throughout conditions, predictions were estimated for 2,500 different random item sets. The mean deviations from the initial predictions for Experiment 1 ($M = .02$, $SD = .02$), Experiment 2 ($M = .03$, $SD = .02$), and Experiment 3 ($M = .03$, $SD = .03$) were small, indicating that the estimated parameters were not idiosyncratic to the random items used to obtain them.

(Manuscript received April 19, 2005;
revision accepted for publication July 1, 2005.)