Memory & Cognition
2007, 35 (3), 432-443

Subtyping as a knowledge preservation strategy in category learning

Lewis Bott and Gregory L. Murphy
New York University, New York, New York

Subtyping occurs when atypical examples are excluded from consideration in judging a category. In three experiments, we investigated whether subtyping can influence category learning. In each experiment, participants learned about a category where most, but not all, of the exemplars corresponded to a theme. The category structure included a *subtyping* dimension, which had one value for theme-congruent exemplars and another for exception exemplars. On the basis of work by Hayes, Foster, and Gadd (2003) and studies on social stereotyping, we hypothesized that this subtyping dimension would enable the participants to discount the exception exemplars, thereby facilitating category learning. Contrary to expectations, we did not find a subtyping effect with traditional category-learning procedures. By introducing the theme prior to learning, however, we observed increased effects on typicality ratings (Experiment 1) and facilitated learning of the category (Experiment 2). Experiment 3 provided direct evidence that introducing the theme prior to learning enhanced the subtyping effect and provided support for a knowledge-gating explanation of subtyping. We conclude that subtyping effects are strongest on already-learned concepts and that subtyping is unlikely to aid in the learning of new concepts, except in particular circumstances.

Learning a new concept is greatly facilitated when prior knowledge can be brought to bear on it (see, e.g., Heit & Bott, 2000; Kaplan & Murphy, 2000; Murphy & Allopenna, 1994; Rehder & Murphy, 2003). One problem with knowledge, however, is that it is sometimes wrong. Even when it is not wrong, it is often rather shallow, not explaining phenomena in very great detail (Keil & Wilson, 2000; Rozenblit & Keil, 2002). One might wonder, then, how real-world knowledge can manage to help us learn anything new. If we have a belief about why birds fly, having to do with wings, then how do we explain turkeys or penguins, which have wings but do not fly? And if we cannot, in fact, make such predictions on the basis of our knowledge, is it any use at all?

Fortunately, research has shown that even if one's prior knowledge does not relate all the features in the concept, it still aids in the learning of the concept (Kaplan & Murphy, 2000). Furthermore, if some of the knowledge is wrong in individual cases, the knowledge still helps people learn, so long as it is generally correct (Murphy & Kaplan, 2000). So, however knowledge influences concept acquisition, it does not require unrealistic levels of perfection to be helpful. The occasional turkey does not prevent us from understanding how birds usually fly and from using that knowledge to learn about new flying animals.

There are other ways in which knowledge may persist even when it is not completely correct. Researchers concerned with social stereotypes have examined how it is that stereotypes can persist in the face of disconfirming group members. As a general rule, people are very reluctant to change their views about social categories, even when there is abundant evidence contradicting them (see Hilton & von Hippel, 1996, for a review). For example, Stephan (1985) demonstrated that negative stereotypes continued to be maintained even after long periods of cooperation with members of the stereotyped group. One of the strategies used to maintain these beliefs in the face of disconfirming information is known as *subtyping* (e.g., Hewstone & Hamberger, 2000; Hewstone, Hassebrauck, Wirth, & Waenke, 2000; Kunda & Oleson, 1995). Subtyping is the process by which group members who disconfirm the stereotype are clustered together to form a subgroup. By segregating such members, the remaining group members can be interpreted as the "real" group, which does, in fact, maintain the stereotype.

The effect of subtyping is to reduce the belief change necessitated by disconfirming examples, as measured by ratings of the stereotypical belief. For example, if you believed that British people are snobbish and then you met a very unsnobbish British subject, you might find a way to isolate this person from the rest of the class. After all, this person is Scottish, and they are not as snooty. Or maybe she is a travel agent, and of course, they have to be friendly as part of their business. As this example illustrates, the subtype is often defined by some other feature that justifies segregating the inconsistent group members. And although such subtyping may be illegitimate in maintaining stereotypes, it is not always unreasonable, since large birds, such as turkeys, often do not fly, for a very good reason. Thus, segregating counterexamples of this sort may

L. Bott, bottla@cardiff.ac.uk

be a reasonable practice if the subtyping feature is, in fact, predictive of a subtype.

This process of using another attribute to explain away counterexamples to stereotypes has been documented in the laboratory. For example, Kunda and Oleson (1995, Experiment 1) studied how the perception of lawyers changed after an encounter with a shy lawyer who violated the stereotype that lawyers are extroverted. In their control condition, participants read a transcript of an interview with a shy lawyer and then rated the extroversion of lawyers in general. These participants rated lawyers as being reliably less extroverted after reading the transcript than did a control group of participants who had not encountered the shy lawyer. In other conditions, the participants were again given the transcript of the interview with the shy lawyer, but they were also given the information that the lawyer worked for a "small firm" or else a "large firm." The interesting finding was that those participants who were given information about the size of the firm failed to change their beliefs; they rated lawyers as being extroverted to the same extent as did those who had not encountered the shy lawyer. Apparently, they reasoned something like, "Well, of course, a lawyer who is part of a small/large firm would very likely be shyer than the usual lawyer," adducing reasons for this particular conclusion. Thus, the availability of a property that could be used to subtype the item allowed the participants to maintain their beliefs about lawyers in general. This sort of reasoning also takes place in other domains, such as evaluating scientific evidence (Chinn & Brewer, 2001).

Subtyping and Category Learning

Research on subtyping in stereotypes has investigated how people maintain a prior belief in the face of contrary evidence. However, it is not so clear to what degree such mechanisms apply in the learning of new facts or categories. Hayes, Foster, and Gadd (2003) examined how school-age children evaluated evidence about a new set of people when subtyping information was available. They used an observation-learning paradigm based on Heit's (1994), in which participants viewed exemplars (e.g., a described child) who were congruent or incongruent with social stereotypes, together with a stereotype-neutral feature (the subtyping feature). For example, one exemplar might contain the following properties: has long hair, wears a dress, has blue eyes (congruent with the gender stereotype that girls have long hair and wear dresses). Another exemplar might contain the following properties: has short hair, wears a dress, has brown eyes (incongruent with the stereotype that boys have short hair but girls wear dresses). In the subtyping condition, the incongruent exemplars were always paired with one subtyping feature (e.g., brown eyes), whereas the knowledge-congruent exemplars always occurred with a different one (blue eyes). In the control condition, the subtyping feature did not correlate with the congruence of the exemplars. After observing these examples, the participants had to judge which features co-occurred most often.[1] The results of the control condition were similar to those in Heit (1994), in that the participants generally selected the knowledge-congruent feature pairings over the knowledge-incongruent feature pairings. But when the subtyping feature covaried with the congruence of the exemplars, the participants selected the congruent feature even more. That is, children who had been trained in the presence of the subtyping covariate were more likely to rely on their prior knowledge than were those who had not. They seemed to be thinking something like, "most of the kids are normal, except for those weird brown-eyed ones," and therefore, they gave less weight to the contradictory examples and claimed that most children did fit the stereotype. Hayes et al. explained their results by suggesting a gating mechanism in the learning system that is sensitive to the degree to which new exemplars fit expectations. When newly encountered exemplars are sufficiently congruent with prior expectations, the gating mechanism allows these exemplars to be incorporated into the category representation. When the exemplars violate expectations, the gating mechanism prevents these exemplars from being incorporated into the category, and it remains unchanged. When a subtyping feature is present, it signals the incongruence of the exception exemplars and indicates that they should not be included in the category representation. Thus, subtyping increases the effect of prior knowledge.

The work of Hayes et al. (2003) is important because it suggests that subtyping could be a phenomenon that occurs in the learning of categories, as well as in the evaluation of members of known categories. Furthermore, their proposed mechanism suggests that subtyping effects could well be found outside the domain of social stereotypes (although their materials were all of this sort). Knowledge effects on category learning have been demonstrated in all sorts of domains, and there is no obvious reason why this gating mechanism should not occur in learning all sorts of categories.

However, before this conclusion can be drawn, we must overcome some of the limitations of the previous work. First, subtyping must be shown in a domain other than social stereotypes. Second, in order to apply to category learning in general, it should be tested in the more usual sort of learning procedure in which participants attempt to classify items and learn the category structure through corrective feedback, which Hayes et al. (2003) did not do. Although classification is only one way to learn categories (Markman & Ross, 2003), it is an extremely important way that is the basis for most models of category learning. Third, materials should be used whose co-occurrence have not already been strongly represented. People already know that children with long hair tend to wear dresses, rather than trousers (or they believe this stereotype, whether or not it is true), but in learning most concepts, people are not learning strongly associated features such as these. Thus, in our experiments, we used materials from past knowledge-based concept acquisition studies that make good sense together but that were not strongly associated in advance of learning the category.

If the subtyping feature can trigger a gate for the application of stereotypical beliefs, it might also facilitate category learning when exception items are present in a training set. Consider a situation in which people are learning a new concept and trying to map their prior knowledge onto the new environment but there are some items for which the knowledge mapping does not seem to apply. If a subtyping feature covaries with the exception items, people should be able to discount the exception items and prevent the knowledge mapping from being destroyed. Without a subtyping feature, the person might never learn the category, because counterexamples would prevent the mapping from forming (see Heit, Briggs, & Bott, 2004, Experiment 3; Murphy & Kaplan, 2000).

Overview of Experiments

We report the results of three experiments in which the effects of subtyping on category learning was investigated. In each case, participants saw exemplars that formed novel vehicle categories. We used categories in which the majority of the exemplars in one category corresponded to a theme—namely, hot-climate vehicles or cold-climate vehicles (themes used by Kaplan & Murphy, 1999, and Murphy & Allopenna, 1994). In addition to the exemplars that conformed to the theme, there were several exception examples that were incongruent with the others. That is, four items in the hot-climate category were cold-climate vehicles, and vice versa. These exceptions corresponded to the shy lawyer in Kunda and Oleson (1995) or the long-haired child who liked playing with toy trucks in Hayes et al. (2003). We then added an additional feature to each exemplar (the vehicle manufacturer) that perfectly predicted the exception items: The standard items all had one feature, *built by the General Vehicles Corporation*, and the exception items had a different feature, *built by Amazing Adventure Vehicles*. This subtyping feature could therefore be used to "explain away" the exceptions, as in Hayes et al.

Pilot experiments. In three experiments, we attempted to find effects of subtyping features on category learning—using both the traditional two-category classification-learning procedure and a one-category design (as in Experiment 1 below). As has just been described, the majority of the exemplars followed the theme of the category, but a minority of them did not. In none of these experiments were we able to find evidence that a subtyping feature aided learning or, indeed, that the participants even noticed the subtyping feature at all. This failure, in contrast to the results of Hayes et al.'s (2003) study and the earlier social-subtyping experiments, was puzzling.

One possible explanation for our failure to find subtyping effects is that in our experiments, the participants were expected to derive the theme linking the examples, whereas elsewhere, researchers used features that were related together through beliefs known prior to the experiment. For example, people's explaining away of the shy lawyer (Kunda & Oleson, 1995) took place within the context of a well-known stereotype that lawyers are not shy. Simply reading that someone is a lawyer likely activates such stereotypes. Hayes et al. (2003) used gender differences that would have been universally known and salient to their subject population. In a category-learning context, however, people had to identify the particular theme of the category, which was not very familiar, in a context in which a number of exemplars were inconsistent with that theme. Perhaps identifying the theme despite exceptions and also noticing that the subtyping feature was correlated with the exceptions (which requires correctly identifying the theme) was just computationally too much for the participants to do.

To make our experiments more similar to those in the social psychology literature, in which stereotypes were already well known prior to the experiment, we informed the participants of the hot/cold climate theme before they saw the vehicles. This would "entrench" the beliefs about the categories, even though the specific features still had to be learned. We believed that the participants would attempt to justify why there were items that did not fit in with the theme and, in the process, discover the covariation of the subtyping feature. That is, in trying to decide why some of the items did not match the stated theme, they would notice the subtyping feature and use that to explain away the discrepancy. In Experiment 3, we directly investigated whether providing the theme in advance of learning was necessary to obtain a subtyping effect.

In Experiment 1, we investigated whether subtyping effects could be observed with our materials when the participants knew the category theme before they saw the examples. We used a typicality-rating task with a single category because this was the design most similar to the original subtyping experiments in the social domain. In earlier experiments, the focus was on a single feature that was atypical (e.g., shyness in a lawyer, or an atypical feature for a correlation in Hayes et al.'s [2003] experiment). However, from the perspective of category learning, an exception item is generally taken to be an exemplar that is actually more similar to a contrast category than it is to its own category (as in nonlinearly separable categories; see, e.g., Smith & Minda, 1998, and references therein). And within most categories, even normal items may have a single unusual property without being considered exceptions (e.g., dining room chairs do not have arms; cardinals have an atypical color). Thus, in studying subtyping in category learning, we focused on exception items that were globally dissimilar from their category.

EXPERIMENT 1

In the first phase of Experiment 1, the participants observed a set of exemplars from a single category. The majority of the exemplars corresponded to a theme (hot-climate vehicles), but there were several exception examples that were cold-climate vehicles. Each exemplar also had an additional feature, the subtyping feature, which referred to the vehicle manufacturer. Two groups of participants completed the task. One group saw exemplars from a category structure where the subtyping features covaried perfectly with the exception items, and the other group saw exemplars where the subtyping features did not have such a covariation. After the study phase, both groups of participants judged the typicality of individual features and pairs of features with respect to the category of vehicles that they had just observed.

Both groups were informed of the theme linking together the exemplars: Each exemplar had "Hot-Climate Dealership" written above it, and the participants were told that the vehicles were sold by a company specializing in hot-climate vehicles.

The effect of a subtyping feature on typicality ratings has often been observed in experiments on social subtyping (see Hewstone & Hamberger, 2000; Hewstone et al., 2000; Kunda & Oleson, 1995). For example, Kunda and Oleson demonstrated that when participants read about a shy lawyer with a subtyping feature, they rated lawyers in general as being more extroverted than when the shy lawyer was presented without the subtyping feature. Such findings suggest that our subtyping group would be less influenced by the exception items than would those in the control condition. However, the category name and instructions informing the participants of the hot-climate vehicle theme made it unlikely that effects would be observed on the hot-climate features themselves, because of ceiling effects. We therefore expected to see differences on the cold-climate features only. Specifically, if the knowledge mapping is preserved by the subtyping feature, participants in the subtyping condition would rate the cold-climate features as less typical than would those in the control group.

Method

Participants. Forty New York University students participated in the experiment for pay or course credit. Twenty were randomly assigned to each condition.

Stimuli and Design. The participants were presented with one category of vehicles consisting of 12 exemplars. Each exemplar consisted of four knowledge dimensions and one subtyping dimension. The knowledge dimensions are shown in the first four rows of Table 1. The left side of the table represents the hot-climate vehicle features, and the right side the cold-climate features. The subtyping dimension was *built by General Vehicles Corporation* versus *built by Amazing Adventure Vehicles* for the standard exemplars feature and the exception exemplars feature, respectively.

Exemplars were constructed according to the left side of Table 2, which corresponds to the hot-climate vehicle category. (The right side, corresponding to cold-climate vehicles, is relevant only for later experiments.) There are 12 rows in the table, each row describing a single exemplar. Each exemplar consisted of five features, four of which were knowledge features, shown under the K columns, and the fifth a subtyping feature taken from either the Subtyping Group column or the Control Group column, depending on the condition. Feature values marked as 1 refer to the hot-climate feature values of the relevant dimension (the left side of Table 1), and those marked 0 to the cold-climate feature values (the right side of Table 1). Standard exemplars are those in which most of the features corresponded to the theme of the category (Exemplars 1–8), and exception items (Exemplars 9–12) are those that contained features from the other category. The subtyping feature covaried with the congruence of the exemplars for the participants in the subtyping group, but not for those in the control group.

Note that although there were twice as many examples that were consistent with the theme than were inconsistent, because of the exception features (0s on the left-hand side) and incongruent items, the typical and atypical features were actually equally frequent in each category.

Procedure. To communicate that the vehicles formed a group of hot-climate vehicles, the participants were told in the instructions that they were all sold by the same dealership, the "Hot-Climate Dealership," and this name was written above each exemplar.

In the first phase of the experiment, the participants viewed the exemplars described on index cards. They were told that "All the vehicles belong to the same category; they are all examples of one type of vehicle." The experimenter instructed them to "learn as much as you can about what kind of vehicles they are and what kind of features they have." They were also told that after 10 min of studying the cards, they would perform a computer exercise based on the examples. They then rated the typicality of features presented in pairs and individually. The participants were instructed that they would now see more examples of vehicles but that these vehicles would have some features missing. They were told to imagine what the missing features might be and to rate, on a scale of 1–9, how typical they thought the vehicles were of the category they had just learned.

Results

We first will consider whether the participants in the subtyping group noticed the correlation between the subtyping feature and the exception items. To do this, we compared the typicality ratings of the features paired with the standard subtyping feature and those paired with the exception subtyping feature. Hot-climate features paired with the standard subtyping feature, such as the pairing *drives in jungles/built by the General Vehicles Corporation*, were judged as more typical than were cold-climate features paired with the standard subtyping feature, such as *drives on glaciers/built by the General Vehicles Corporation* ($M = 7.08$, $SD = 0.98$, vs. $M = 5.57$, $SD = 1.48$); yet this pattern was reversed when the features were paired with the exception feature ($M = 3.75$, $SD = 2.32$, vs. $M = 4.43$, $SD = 2.50$). The interaction of climate and subtyping feature was reliable [$F(1,39) = 17.85$, $p < .001$]. Thus, the participants in the subtyping group had learned that the hot-climate features typically occurred with the standard subtyping feature, and not with the exception subtyping feature.

The comparison of interest between the subtyping and the control group was on the (atypical) cold-climate features. We combined the single- and double-feature scores, weighted by the number of trials in each cell, to obtain a single figure for each participant, representing their typicality ratings for cold-climate features. The mean typicality ratings were 4.86 ($SD = 1.03$) for the subtyping group and 5.71 ($SD = 1.45$) for the control group, which was a reliable difference [$t(38) = 2.13$, $p < .05$]. Thus, we can conclude that the participants in the subtyping condition had exaggerated effects of prior knowledge, in compari-

Table 1
Knowledge Dimensions Used in the Experiments

Dimension Number	Hot-Climate Features	Cold-Climate Features
1	Drives in jungles	Drives on glaciers
2	Used on safaris	Used on mountains
3	Made in Africa	Made in Norway
4	Lightly insulated	Heavily insulated
5	Green	White
6	Has wheels	Has treads

Note—Dimensions 1–4 were used in Experiments 1 and 2, and Dimensions 1–6 were used in Experiment 3.

Table 2
Abstract Category Structures for Experiments 1, 2, and 3

Hot-Climate Category							Cold-Climate Category						
	Knowledge Dimension				Subtyping Dimension			Knowledge Dimension				Subtyping Dimension	
Exemplar Number	K1	K2	K3	K4	Subtyping Group	Control Group	Exemplar Number	K1	K2	K3	K4	Subtyping Group	Control Group
1	0	1	1	1	1	0	13	1	0	0	0	1	0
2	1	0	1	1	1	0	14	0	1	0	0	1	0
3	1	1	0	1	1	0	15	0	0	1	0	1	0
4	1	1	1	0	1	0	16	0	0	0	1	1	0
5	0	1	1	1	1	1	17	1	0	0	0	1	1
6	1	0	1	1	1	1	18	0	1	0	0	1	1
7	1	1	0	1	1	1	19	0	0	1	0	1	1
8	1	1	1	0	1	1	20	0	0	0	1	1	1
9	0	0	0	0	0	1	21	1	1	1	1	0	1
10	0	0	0	0	0	1	22	1	1	1	1	0	1
11	0	0	0	0	0	1	23	1	1	1	1	0	1
12	0	0	0	0	0	1	24	1	1	1	1	0	1

Note—Exemplars 1–12 form the hot-climate category, and Exemplars 13–24 form the cold-climate category. Exemplars 9–12 and 21–24 are exception exemplars.

son with the control group, rating the cold-climate features as more atypical. The hot-climate features were rated uniformly high in the two groups, as was expected (*M*s = 6.88 and 6.60 for the subtyping and control groups, respectively), presumably because both groups were instructed that the category represented hot-climate vehicles.

Discussion

Experiment 1 showed that subtyping effects can be observed using novel categories, when the category has a theme running through it and the participants are aware of that theme before observing the examples. This result adds to knowledge of the subtyping phenomenon, because it demonstrates that subtyping is not restricted to traditional social stereotypes but applies to any suitably themed category. It also extends subtyping to cases in which an entire exemplar (not just a single property) is atypical.

We have described pilot experiments that did not show evidence of subtyping, suggesting that learners must know the category theme in advance in order to take advantage of the subtyping feature. Finding that people could use the subtyping feature in the present experiment supports this proposal. We explicitly tested this notion in Experiment 3. Having now established conditions in which participants notice the correlation between the subtyping feature and the exception items, we next return to the central question of the article—namely, whether subtyping can facilitate the learning of a concept and, if so, how.

EXPERIMENT 2

Experiment 2 was a category-learning experiment in which the participants learned to discriminate two categories of vehicles involving standard exemplars and exception items. One category consisted of mainly hot-climate vehicles, as in Experiment 1, whereas the other category consisted of mainly cold-climate vehicles. Exception exemplars were those that consisted of features from the category other than that to which they were assigned. One group of participants was taught a category structure involving a subtyping feature that covaried with the presence of the exception items, and one group received a control category structure. Having observed subtyping effects in Experiment 1, we employed a similar category structure in this experiment, and we also informed the participants of the theme characterizing the two vehicle categories before learning commenced. After learning the exemplars to a criterion, the participants proceeded on to a feature-testing phase in which they were tested on individual features.

The participants in the subtyping group might learn the category structure more quickly than those in the control group because they could use the subtyping feature to decide whether to apply the hot-/cold-climate mapping to that exemplar. Those in the control group would not have such a gating strategy open to them and would, therefore, have to abandon their use of the knowledge mapping in the face of contradictory examples. Examples that strongly violate a category's theme make learning much harder (Murphy & Kaplan, 2000). Furthermore, we would expect differences between the groups for the individual-feature-testing phase. If the participants in the control group abandoned the prior knowledge mapping, they should assign individual features to categories arbitrarily, because any given feature value occurred equally often in both the hot-climate category and the cold-climate category (see the Method section for more details). If the participants maintained the prior knowledge mapping, they should place the hot- and cold-climate features into the hot- and cold-climate categories, as appropriate.

Method

Participants. Twenty-eight New York University students participated in the experiment for pay or course credit. Fourteen participants were randomly assigned to each of the two conditions.

Stimuli and Design. The participants saw exemplars from two categories, the hot-climate vehicles and the cold-climate vehicles. The category structure is shown in Table 2, where the left side corresponds to the hot-climate vehicle category and the right to the cold-climate vehicle category. Each exemplar was constructed from four knowledge features (Dimensions 1–4 in Table 1) and one subtyping feature. As in the previous experiment, there were standard exemplars that conformed to the theme and exception items that did not (Exemplars 9–12 and 21–24, respectively, in Table 2). The subtyping feature covaried with the congruence of the exemplars for those participants who were placed in the subtyping group, but not for those in the control group, as is shown in Table 2.

Procedure. The participants were told that they would be learning about vehicles sold by two different dealerships, Dealership A and Dealership B, and that they would have to learn to classify the vehicles by paying attention to the feedback. They were also explicitly told that "Dealership A sells mostly hot-climate vehicles while Dealership B sells mostly cold-climate vehicles. Note that not every vehicle sold by the company follows this hot-/cold-climate distinction, but on the whole, this generalization holds." Labels saying "Hot" and "Cold" were also placed on the response keys below the A and B category labels.

Exemplars were presented as written descriptions on a computer screen, with one exemplar presented per screen. The participants read the description of the exemplar and pressed a key corresponding to a category. They then received feedback indicating whether they were correct and what the true classification should have been. Learning proceeded in blocks consisting of the presentation of all the exemplars shown in Table 2, in a random order. If a participant had succeeded in classifying all of the exemplars of a block correctly, he or she entered the individual-feature-testing phase; if not, learning continued for up to 16 blocks.

In the feature-testing phase, the features were presented individually on the screen, and the participants pressed a key to indicate whether the feature was most likely to be from Category A or B. Feedback was not provided. Each participant classified each feature twice, making a total of 20 test trials (2 × [8 knowledge features + 2 subtyping features]). At test, the participants were told that they would be seeing some new vehicles but that they would be able to see only one feature from these vehicles. They were instructed to imagine what the other features might be and to decide which of the two categories the new vehicle would be most likely to belong to.

Results

Learning phase. Learning was easier with the subtyping structure. In the subtyping group, 9 out of 14 participants learned the category structure within the 16-block limit, whereas only 2 out of 14 participants reached this criterion in the control group. This difference is significant by Fisher's exact test ($p < .05$). Similarly, with those participants who did not complete within 16 blocks receiving a generous score of 17, the participants in the subtyping group required reliably fewer blocks to complete the learning phase ($M = 11.0$, $SD = 1.5$) than did those in the control group ($M = 16.4$, $SD = 5.9$) [$t(26) = 3.4$, $p < .005$]. Thus, the participants were able to make use of the subtyping feature in learning the concept.

Figure 1 displays the performance of the two groups on the standard and exception exemplars as a function of learning block. To avoid empty cells in later blocks, we assigned the participants a score of 1.0 after they had reached criterion. For example, Subject 14 reached criterion on Block 2; hence, he was assigned an accuracy score of 1.0 for both the standard and the exception exemplars for Blocks 3–16. The upper two lines in Figure 1 show that for the standard exemplars, both groups performed accurately early in learning (recall that they were provided with the theme at the outset) and that the subtyping group achieved higher scores in later blocks. This is demonstrated by an interaction between the effect of learning block and category structure (subtyping vs. control) [$F(15,390) = 1.84$, $MS_e = 0.014$, $p < .05$; standard exemplars only]. More interesting is the pattern of responses to the exception exemplars, which differed across the two groups [$F(15,390) = 10.25$, $MS_e = 0.049$, $p < .001$; exception exemplars only]. Although both groups started out assigning the exceptions according to the theme (i.e., incorrectly), the subtyping group steadily improved its performance, presumably reflecting the discovery of the subtyping structure that identified the exception items. The control group did not show this steady increase, improving only after 12 blocks, probably as a result of memorizing the exceptions.

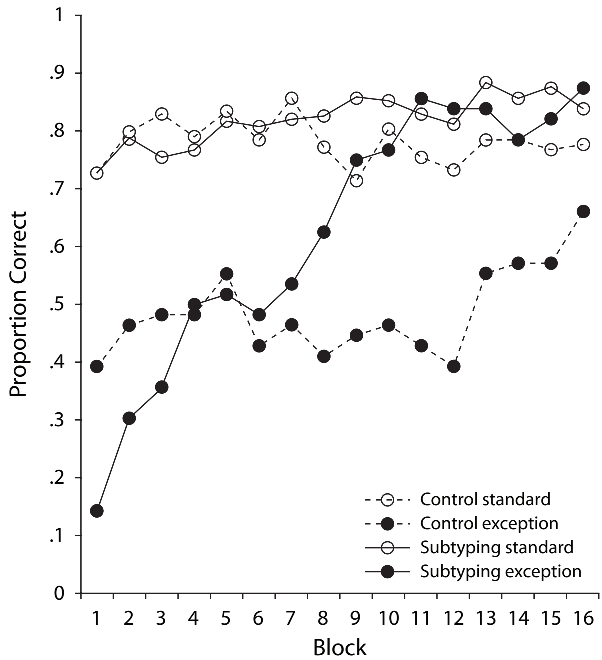

Figure 1. Proportion correct classifications during learning of the standard and exception exemplars when participants learned the subtyping structure (*subtyping* condition) and when they learned the control structure (*control* condition) in Experiment 2.

Test phase. Most participants responded very accurately to the features in the feature-testing phase, regardless of whether they were in the subtyping group (M proportion correct = .81, $SD = .19$) or the control group ($M = .92$, $SD = .11$). Although this difference approaches statistical significance [$t(26) = 1.95$, $p = .062$], it is somewhat difficult to interpret, given that the control group had five more blocks of exposure to the features, on average. It is possible that the control group relied more on feature learning than on the theme, because the subtyping feature was not available to explain away the theme violations.

Discussion

The participants in the subtyping condition learned the category structure more quickly than did those in the control group. This is an important result because it demonstrates that the subtyping phenomenon applies to supervised category learning, as well as to the process of classification. Furthermore, that participants can use the subtyping feature in a category-learning task entails that they are sensitive to the feature correlations within the exemplars, which is not generally found in supervised learning (Chin-Parker & Ross, 2002).

The subtyping group's performance on the exception exemplars was slightly worse early in learning, but then, as more blocks were experienced, they learned these and the standard exemplars better than did the control group, on average, so that more participants reached the learning criterion. What type of mechanism was responsible for the facilitation in learning? We will consider two possibilities. The first is that the subtyping feature acts to direct attention to the different types of items. Initially, participants attend equally to all exemplars, but then, as more exemplars are observed, they realize that the exception exemplars require more resources. Focusing attention on the difficult items results in more efficient learning. An alternative is that the subtyping feature acts as a gate to allow the knowledge mapping to be applied in some situations and not in others (Hayes et al., 2003). Early in learning, participants do not abandon the theme because they feel they can explain away the exception exemplars. When there is a subtyping feature, this attempt is successful, but when there is not, the knowledge mapping may have to be abandoned.

These two possibilities can be distinguished by investigating how people use the subtyping feature when generalizing beyond the features they see in the experiment. If participants use the subtyping feature to activate a knowledge gate, their classification of novel features should be affected by the presence of the subtyping feature, just as it is for the classification of old features. With one subtyping feature, they ought to assign novel features to the thematically consistent category; with the other, they should assign it to the "wrong" category. If the subtyping feature acts merely to mark certain exemplars as unusual during learning, it will not provide any specific information about how to classify a new feature. Our next experiment included the novel features necessary to distinguish between these two mechanisms.

Another goal of Experiment 3 was to determine the effect of providing the theme in advance of learning. We suggested that one of the reasons we obtained subtyping effects in Experiments 1 and 2, but not in our pilot experiments, was that the participants were informed of the theme in the former experiments, but not in the latter. The next experiment directly tested this possibility by comparing the performance of participants who did or did not receive the theme in advance.

EXPERIMENT 3

The participants were divided into two groups. One group received instructions relating the exemplars to the climate theme of the vehicles, as in Experiment 2, and the other group received neutral category-learning instructions. Apart from the instructional manipulation, the learning phase of the experiment was identical to that in the subtyping condition of Experiment 2. (The control condition would not help to determine whether subtyping was facilitated with advance knowledge of the theme and so was omitted.) A further difference between Experiments 2 and 3 occurred in the feature-testing phase. In Experiment 3, the participants were tested not only on individual features, but also on pairs of features, some of which the participants had not seen before. This enabled us to assess generalization behavior in relation to the subtyping feature.

We hypothesized that knowing the theme in advance of learning would facilitate the subtyping effect. Under the themed instructions, the participants would seek to understand why the exception exemplar was not in the expected category, thereby noticing the covariation between the subtyping feature and the exception exemplars. Without knowing the theme in advance, the participants would find it more difficult to notice that subtyping features predicted whether an item was theme consistent or inconsistent. An enhanced subtyping effect might manifest itself in several ways. First, participants who used the subtyping feature might find the category structure easier to learn, as we observed in the previous experiment. Second, participants who used the subtyping feature should vary their responses to the standard features as a function of the accompanying subtyping feature. For example, if the feature *drives in jungles* is classified as an A vehicle in the presence of the subtyping feature *built by General Vehicles Corporation*, then, if the participant is aware of the significance of the subtyping structure, classification should change when the same feature is accompanied by the *built by Amazing Adventure Vehicles* subtyping feature. If knowing the theme in advance encourages subtyping, we would expect these effects to be more pronounced in the themed condition.

Method

Participants. Thirty New York University students participated in the experiment for pay or course credit. Fifteen were randomly assigned to each condition.

Stimuli and Design. The formal category structure was identical to the subtyping condition in Experiment 2. However, we added two features per category for use in the testing phase. The knowledge features used in this experiment were the first six dimensions shown in Table 1. The assignment of dimensions to presentation type was rotated across participants, so that each participant saw a different combination of novel and presented dimensions.

The participants in the themed condition were given instructions relating the categories to the hot-climate and cold-climate theme, exactly as in Experiment 2. Those in the neutral condition were given standard category-learning instructions (i.e., no theme). All other aspects of the learning phase were identical to those in Experiment 2.

During the feature-testing phase, the participants viewed single features and pairs of features on the computer monitor and assigned them to categories. Each of the 14 features (7 dimensions) was presented twice individually and twice with each other feature as part of a pair, one feature below the other. Thus, there were a total of 28 single-feature trials and 168 paired-feature trials (no features were paired with their binary opposites). The instructions for the test phase were identical to those in Experiment 2, except that the participants were now told that they would be able to see only some of the features from the new vehicles.

Results

Learning phase. We first consider the effects of knowing the theme in advance of learning. We suggested that the theme should encourage subtyping, thereby facilitating learning and altering response patterns in the feature-testing phase. In the learning phase, more participants learned in the themed condition than in the neutral condition (7 out of 15 vs. 3 out of 15), although the difference was not reliable (Fisher's exact test, $p = .25$), nor was it reliable when we compared the trials with criterion in the two groups (conservatively giving 17 for those who failed to learn) [$M = 12.6$, $SD = 5.72$, vs. $M = 12.8$, $SD = 4.62$; $t(29) = 1.23$, $p = .24$]. Those participants who learned in the themed condition required 7.3 ($SD = 4.19$) blocks to reach criterion, on average, and those in the neutral condition required 6.0 ($SD = 2$) blocks.

We also analyzed the relative learning rate for the standard and the exception exemplars, as in the previous experiment. Figure 2 shows block-by-block learning data for the theme and neutral conditions. We again assigned an accuracy score of 1.0 to all blocks after a participant reached criterion. We first consider performance by those participants who were given the theme in advance. Recall that this condition was a direct replication of Experiment 2's subtyping condition; performance for these participants was, therefore, very similar: They began with high accuracy on the standard exemplars and very low accuracy for the exception exemplars. The difference between the two types of exemplars diminished as learning continued, with performance on the exception exemplars becoming more accurate, as indicated by a reliable interaction between learning block and type of exemplar [standard vs. exception; $F(15,210) = 5.53$, $MS_e = 0.0283$, $p < .001$]. Next, consider the neutral group. Its accuracy improved as more blocks were experienced [$F(15,210) = 3.44$, $MS_e = 0.042$, $p < .001$], and the participants found the exception exemplars more difficult than the standard exemplars [$F(1,14) = 6.19$, $MS_e = 0.29$, $p < .05$]. However, there was no change in relative difficulty between the standard and the exception items as more blocks were observed [$F(15,210) = 1.32$, $MS_e = 0.030$, $p = .20$], in striking contrast to the participants who received the theme in advance. Thus, although the participants in the neutral group were able to extract the theme to some extent, they appeared unable to assimilate the exception items as the experiment continued, unlike those in the theme condition. This is confirmed by a reliable three-way interaction between block, exemplar type, and theme condition [$F(15,420) = 2.10$, $MS_e = 0.029$, $p < .005$]. The theme manipulation therefore affected learning, despite there being no reliable difference in the number of participants who reached criterion.

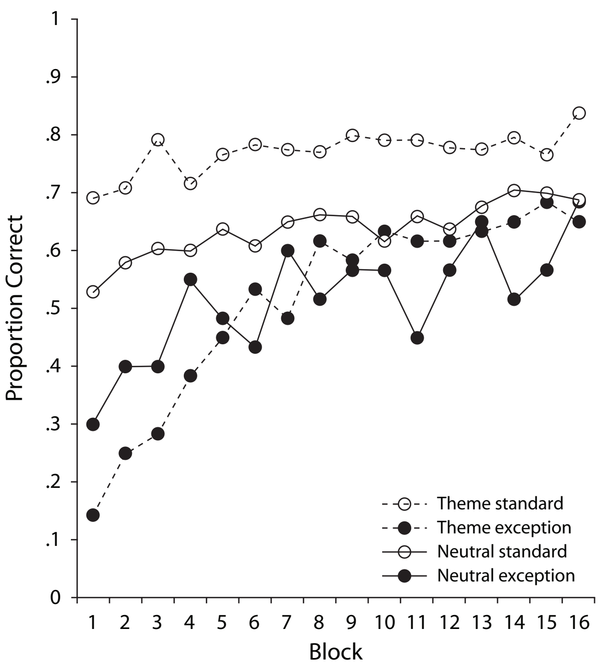

Figure 2. Proportions of correct classifications during learning of the standard and exception exemplars when the theme was provided first (*theme* condition) or not (*neutral* condition) in Experiment 3.

Test phase. The participants classified individual features and feature pairs. The most relevant results were those that involved pairs where one of the two features was a subtyping feature—for example, *drives in jungles*/*built by Amazing Adventure Vehicles*. If the participants were using the subtyping feature, their response to hot-climate features paired with the standard subtyping feature should be different from that to hot-climate features paired with the exception subtyping features. The same pattern should hold for the cold-climate features. The proportion of hot-climate category responses for all the participants is shown in Table 3 as a function of the presented theme and the pairing of the subtyping feature. Thus, if the participants used the subtyping feature, they should have high scores for the hot-climate/standard subtyping feature pairing but low scores for the hot-climate/exception feature pairing, and vice versa for the cold-climate features. This pattern was indeed found in the themed instructions condition (first row of data in Table 3), but not in the neutral instructions condition (second row in Table 3).

To determine whether this pattern was reliable, we calculated a subtyping score for each participant on the basis of the extent to which they varied their classification of the vehicle features in the presence of the different subtyping features. We defined this subtyping index as (HS0 − HS1) − (CS0 − CS1), where HS0 indicates a hot-climate feature paired with the standard subtyping feature, HS1 indicates a hot-climate feature paired with the exception subtyping feature, and CS0 and CS1 refer to the cold-climate feature equivalents. The dependent measure in each case was the proportion of hot-climate feature responses. The maximum score on this scale is, therefore, +2, indicating that the participant varied his or her responses perfectly in line with the subtyping pattern shown in the category. A score of 0 indicates no effect of the subtyping feature, and a score of −2 indicates a subtyping pattern reversed with respect to the category structure. By this index, the participants in the themed condition displayed more of a subtyping effect than did those in the neutral condition, who, on average, ignored the subtyping feature [$M = 0.63$, $SD = 0.81$, vs. $M = 0.02$, $SD = 0.53$; $t(28) = 2.47$, $p < .05$]. The difference between the two groups was reliable for features presented during learn-

Table 3
Mean Proportions of Hot-Category Responses to Each Type of Feature Pairing for Experiment 3 (With Standard Deviations)

	Feature Pairing							
	Standard/Hot		Exception/Hot		Standard/Cold		Exception/Cold	
Instructions	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Themed	.78	.26	.49	.37	.20	.17	.55	.33
Neutral	.50	.23	.53	.31	.42	.25	.47	.21

Note—The first term in each column label refers to the value of the subtyping feature (standard or exception), and the second to the type of knowledge feature (hot or cold).

ing [$M = 0.63$, $SD = 0.84$, vs. $M = -0.01$, $SD = 0.56$, $t(28) = 2.43$, $p < .05$] and marginally so for the novel features [$M = 0.65$, $SD = 0.83$, vs. $M = 0.07$, $SD = 0.78$; $t(28) = 1.98$, $p = .057$]. Furthermore, the subtyping effect was reliably different from zero in the themed condition for both the presented and the unpresented features [$t(14) = 2.89$, $p < .05$; $t(14) = 3.04$, $p < .01$], whereas neither was reliable in the neutral condition [$t(14)$s < 1]. These results confirm that knowing the theme in advance facilitated use of the subtyping feature.

We now will turn to the second goal of this experiment—namely, to address the mechanisms underlying subtyping. Since we were concerned with how the subtyping feature was used to learn the categories, we included only those who actually learned in subsequent analyses. This meant that 10 participants were included in total, 7 from the themed condition and 3 from the neutral condition (there were insufficient numbers to analyze the learners from the two groups separately). For the features presented during the learning phase, the majority of the participants had high scores on the subtyping index, with $M = 1.11$, $SD = 0.82$, which was reliably different from zero [$t(9) = 4.27$, $p < .005$]. This indicates that these participants had learned the covariation between the subtyping feature and the hot-climate–cold-climate mapping, as we would expect if they were using the subtyping feature to help them learn the category structure. Identical results were observed for the novel items [$M = 1.13$, $SD = 0.84$; $t(9) = 4.22$, $p < .005$], confirming that the subtyping feature was not was simply an indication that the item should receive special attention but that it was used as a way of gating the use of prior knowledge.

Discussion

We found that the participants who received the theme in advance of learning used the subtyping features to change their classification of properties in the feature test, showing that they had learned the subtyping structure. Those who did not receive the theme in advance did not generally use the subtyping feature. Thus, the experiment verified our conclusion from the pilot experiments that learners find it difficult to acquire a subtyping structure for a novel category unless the theme is known in advance.

We also found that the participants varied their responses to novel features, depending on the value of the subtyping feature. This result illustrates that the participants were not just memorizing feature pairs or examples but were using the subtyping feature to determine when they should apply their prior knowledge mapping in this environment.

Our analysis of the performance on the standard and exception exemplars during learning again revealed interesting results. Both groups of participants displayed lower accuracy for the exception exemplars than for the standard exemplars, indicating that even the participants in the neutral condition extracted the theme to some extent but that only those in the themed condition were able to find a way of applying the knowledge mapping to the standard exemplars, and not to the exception exemplars. Hence, the difference between the two types of exemplars diminished as more blocks were experienced for the themed group, but not for the neutral group. This is further evidence that providing the theme in advance of learning changed what the participants learned about the subtyping structure, even if it did not result in a reliable increase in reaching criterion (although, in fact, twice as many participants in the themed group did reach criterion).

GENERAL DISCUSSION

The goals of this project were to establish whether the presence of a neutral feature covarying with exception exemplars could facilitate category learning and, if so, how. Our results demonstrate that under appropriate conditions, a subtyping feature can encourage the preservation of knowledge mappings that, ultimately, lead to benefits in category learning. Furthermore, we found evidence that the subtyping dimension was acting as a trigger to gate the application of prior knowledge, as was suggested by Hayes et al. (2003), and not simply highlighting the exception examples for extra attention during learning.

One of the surprising findings from this study was that prior presentation of the theme greatly enhanced the subtyping effect. Indeed, we failed to observe any subtyping effects in pilot experiments in which we did not provide the theme prior to learning. Our explanation for this finding is that informing the participants of the theme linking together the examples made it easier for them to identify the exception items as such and, consequently, to search out reasons for why these unusual exemplars might be in this category in the first place. This search led to discovering the subtyping feature if it was present. This account suggests that under the traditional laboratory conditions of category acquisition, people will rarely invoke a subtyping strategy to deal with unusual exemplars, for the simple reason that they would not notice the necessary correlations. This is in keeping with the conclusion of Murphy and Wisniewski (1989) and Chin-Parker and Ross (2002) that participants do not generally learn within-category feature correlations during supervised category learning. Subtyping effects, however, are precisely within-category feature correlations. Interestingly, our results suggest that if these correlations are related to exemplars that conflict with prior beliefs, people will indeed notice and use them (much as Murphy & Wisniewski concluded that within-category correlations linked to prior knowledge were learned).

Despite our own evidence that subtyping does not arise without providing a theme in advance, Lewandowsky, Kalish, and colleagues (Kalish, Lewandowsky, & Kruschke, 2004; Lewandowsky, Kalish, & Ngang, 2002; Yang & Lewandowsky, 2003, 2004) found subtyping-like effects with stimuli that had no theme running through the categories. These researchers have shown that participants simplify complex learning tasks by acquiring independent parcels of knowledge and that this partitioning is helped when a nondiagnostic *context* variable is included with the exemplars. For example, Yang and Lewandowsky (2003) conducted category-learning experiments in which participants learned about two different types of fish, defined on two continuous dimensions. The classification boundary was complex, consisting of a pair of linear boundaries that joined at a vertex in the middle of the space. The predictor variables were accompanied by a binary context variable that divided the space up into two linearly defined boundaries. The context variable was nondiagnostic of the category, much like our subtyping feature, and the participants were able to use the variable to help them divide up the nonlinear mapping into two linear mappings, applying a different linear mapping under different values of the context variable. Why, then, were the participants in Yang and Lewandowsky's (2003) study able to use the context variable without any kind of theme, whereas those in our experiments without themes could not? We cannot provide a definitive answer to this question, but we can point out that there were many concrete differences between our experiments. For example, the tasks differed in the type of features used to describe the exemplars (continuous numerical vs. binary conceptual), the number of exemplars (40 vs. 24), and the category structure (distributional vs. family resemblance plus exception). Furthermore, the context variables in their experiment were correlated with a predictor variable and served to divide the stimulus space in half, whereas our subtyping feature occurred only with the exception items and was uncorrelated with feature values. Finally, their task was particularly difficult (e.g., control condition participants achieved a mean score of only 68% correct at the end of training in Experiment 1), perhaps thereby forcing the participants to look at the internal relationships of the features and discover the effect of the context variable. Future work will have to be done to investigate which of the many differences between the tasks determines when context/subtyping variables of this sort are learned.

Implications for Models of Category Learning

The participants acquired the category structure by learning about the general mapping and the exceptions to this general mapping. When the subtyping dimension correlated with these exceptions, the participants acquired this additional information and found the category structure easier to learn. The challenge facing any model of category learning is to explain why participants would ever learn this correlational information, given that it is perfectly possible to acquire a solution without it, and why providing participants with the theme of the category before learning encourages the subtyping strategy.

One possibility is that the participants merely used a simple associative-learning strategy and that the error surface of our category structure encouraged the participants to find the subtyping solution; perhaps the solution that exists without using the subtyping dimension involves overcoming more local minima, for example. If this were the case, ALCOVE (Kruschke, 1992, 1996) or any other error-driven learning model would predict that the subtyping structure would be learned before the control structure. However, this account fails to explain why we observed subtyping only when the participants were provided with the theme prior to learning. Why should there be an error-driven incentive to attend to the subtyping dimension when the theme is presented, but not in cases in which the theme is absent?

One response would be to argue that the set of initial conditions changes with the introduction of the theme, so that some dimensions are emphasized over others, and the subtyping solution is easier to find with the altered starting state. For example, setting an initially high attention weight on the subtyping dimension, as compared with the other dimensions, would encourage ALCOVE to find a subtyping solution. However, the participants were not instructed to attend to the subtyping feature: They were told that there was a hot-climate theme that prevailed among the vehicles, and the subtyping feature was not semantically related to climate.

The main problem with such models is that there is no way of explicitly incorporating the theme, or rule, into the model before category learning. Thus, it is difficult for them to explain the effects of providing participants with such knowledge. This suggests that dual-component models, such as ATRIUM (Erickson & Kruschke, 1998), BAYWATCH (Heit & Bott, 2000), or KRES (Rehder & Murphy, 2003), might do better at providing an explanation for our subtyping results. These models can represent mappings explicitly, before or after learning. For example, BAYWATCH could represent the participants' knowledge of the theme before learning by assuming a known concept of hot-climate vehicles that overlaps to some degree with the new vehicle category, Category A.

Could such models explain why presenting the theme facilitated use of the subtyping feature? First, note that this mapping facilitates learning of the normal exemplars, because extra activity propagates to the correct category node, via the prior knowledge node. On the other hand, the mapping harms learning of the exception exemplars because this extra activity now propagates to the incorrect category nodes. Because this helpful and harmful activity is perfectly correlated with the subtyping dimension, there is an error-driven incentive to use this dimension to gate the mapping. If this mapping were not in place, there would not be as much of an incentive to use the subtyping dimension.[2] Put more generally, maintaining prior knowledge introduces constraints on the range of allowable solutions to the category-learning problem—constraints that, in this case, make it more likely that the subtyping solution will be among those found by the model. A further advantage of a model such as BAYWATCH is that it could likely reproduce the findings concerning the classification of novel features (Experiment 3). If the model learns not to activate the prior knowledge nodes when the subtyping feature is present during training, the prior knowledge nodes will not become activated when the exception feature is presented in test, even if it is accompanied by novel features. Hence, novel features will be activated according to the theme if they are presented with the standard subtyping feature, but not otherwise.

Other category-learning models have also looked at how people might learn "exception" items, although not for the category structures tested here. For example, RULEX (Nosofsky, Palmeri, & McKinley, 1994) is a rule-plus-exception model of category learning. It attempts to learn simple category rules, adding exception items to the rule if the simple rule is not sufficient. RULEX does not include any prior knowledge component, so it could not account for all the results we have reported here. It is an interesting question, however, whether it would learn the subtyping structure more easily than it would learn the control structure. Although the structure in Experiment 3 (see Table 2) has two-thirds typical items and one-third exception items, the typical and exception *features* are actually equally frequent in each category (because of crossover features). Thus, it might be quite difficult for any system that attempts to learn a feature-by-feature rule to acquire these categories, since each feature on its own is nondiagnostic and it is only when the exemplar is evaluated as a whole (as consistent or inconsistent with the theme) that the family resemblance (and subtyping) structure is evident.

There are many models of category learning, and we cannot test them all against our structures. What seems clear is that no model can account for the present data without representing prior knowledge in some form, because it cannot explain why the subtyping structure is acquired when the theme is presented, but not when it is unknown.

Conclusion

Our study has shown that the subtyping phenomenon is not restricted to exemplar evaluation or to social stereotypes but can be observed in a category-learning task using theme-based stimuli. We also found evidence that the subtyping feature acts as a gate to isolate the knowledge-mapping mechanism from counterexamples, as Hayes et al. (2003) proposed, thus adding to our understanding of the mechanisms underlying the subtyping phenomenon. Surprisingly, we were able to demonstrate subtyping effects only when the theme was known to participants beforehand, as opposed to the usual knowledge-based category-learning situation in which participants acquire the theme during learning. The latter result suggests that although subtyping might be very important for *maintaining* stereotypes and category themes, it is unlikely to be a highly prevalent strategy in *acquiring* new stereotypes and categories.

Our results also have implications for error-driven models of category learning. By finding that participants use the subtyping dimension when they are provided with the theme, but not otherwise, we add to the evidence that models need to be augmented with knowledge to explain how people learn categories.

AUTHOR NOTE

Support for this work came from NIMH Grant MH41704. The authors are grateful to Robyn Kim for collecting data and to Evan Heit and two reviewers for helpful comments on a draft. Correspondence concerning this article should be addressed to L. Bott, School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff CF10 3AT, Wales (e-mail: bottla@cardiff.ac.uk).

REFERENCES

Chinn, C. A., & Brewer, W. F. (2001). Models of data: A theory of how people evaluate data. *Cognition & Instruction*, **19**, 323-393.

Chin-Parker, S., & Ross, B. H. (2002). The effect of category learning on sensitivity to within-category correlations. *Memory & Cognition*, **30**, 353-362.

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

Hayes, B. K., Foster, K., & Gadd, N. (2003). Prior knowledge and subtyping effects in children's category learning. *Cognition*, **88**, 171-199.

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., & 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.

Hewstone, M., & Hamberger, J. (2000). Perceived variability and stereotype change. *Journal of Experimental Social Psychology*, **36**, 103-124.

Hewstone, M., Hassebrauck, M., Wirth, A., & Waenke, M. (2000). Pattern of disconfirming information and processing instructions as determinants of stereotype change. *British Journal of Social Psychology*, **39**, 399-411.

Hilton, J. L., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, **47**, 237-271.

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. (1999). The acquisition of category structure in unsupervised learning. *Memory & Cognition*, **27**, 699-712.

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

Keil, F. C., & Wilson, R. A. (Eds.) (2000). *Explanation and cognition*. Cambridge, MA: MIT Press.

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

Kruschke, J. K. (1996). Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 3-26.

Kunda, Z., & Oleson, K. C. (1995). Maintaining stereotypes in the face of disconfirmation: Constructing grounds for subtyping deviants. *Journal of Personality & Social Psychology*, **68**, 565-579.

Lewandowsky, S., Kalish, M., & Ngang, S. K. (2002). Simplified learning in complex situations: Knowledge partitioning in function learning. *Journal of Experimental Psychology: General*, **131**, 163-193.

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, **129**, 592-613.

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., & Kaplan, A. S. (2000). Feature distribution and background knowledge in category learning. *Quarterly Journal of Experimental Psychology*, **53A**, 962-982.

Murphy, G. L., & Wisniewski, E. J. (1989). Feature correlations in conceptual representations. In G. Tiberghien (Ed.), *Advances in cognitive science: Vol. 2. Theory and applications* (pp. 23-45). Chichester, U.K.: Ellis Horwood.

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

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

Rozenblit, L., & Keil, F. [C.] (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, **26**, 521-562.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge, MA: MIT Press.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1411-1436.

Stephan, W. G. (1985). Intergroup relations. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., Vol. 2, pp. 599-658). New York: Random House.

Yang, L.-X., & Lewandowsky, S. (2003). Context-gated knowledge partitioning in categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 663-679.

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

NOTES

1. In Heit's (1994) original study, participants estimated the frequency of co-occurrence of different features. However, since Hayes et al.'s (2003) participants were children, they simply made binary judgments, and the proportion of children choosing a feature was assumed to indicate the strength of the relation. For example, they might judge whether a child with long hair from the observed school was more likely to wear trousers or a dress.

2. Note that as the mapping is of the XOR type, because the subtyping features are associated to both categories, the solution will require hidden units in order to be learned (see Rumelhart, McClelland, & the PDP Research Group, 1986). Such features are not in the version of BAYWATCH reported by Heit and Bott (2000) but could easily be implemented.

(Manuscript received October 12, 2005;
revision accepted for publication January 4, 2006.)